DOCUMENT RESUME

ED 233 325 CS 007 269

AUTHOR Singer, Harry; And Others
TITLE Testing Achievement in Basic Skills. Volume I: Executive Summary and Research Reports. Final Report.
INSTITUTION California Univ., Riverside. School of Education.
PUB DATE 82
NOTE 222p.; Project funded by the California Policy Seminar. For related document see CS 007 270.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS *Academic Standards; Achievement Tests; *Basic Skills; Educational Research; Educational Strategies; Educational Testing; Elementary Secondary Education; Program Evaluation; Reading Instruction; Standardized Tests; *State Programs; *Test Construction; *Testing; *Testing Problems; Testing Programs; Test Reviews

ABSTRACT

This comprehensive evaluation of basic skills testing in California's elementary and secondary schools is divided into an executive summary and a collection of research reports. The summary evaluates current testing conducted by the state, school districts, and classroom teachers; describes experimental studies on new ways of testing and reporting test information; and makes recommendations on how to improve the testing program in California. The second section, a collection of research reports used in the evaluation, investigates a number of concerns in state, school district, and classroom testing. The topics covered include (1) the evaluation and testing of reading, (2) California's 1981 reallocation of resources to meet minimal competency requirements, (3) attitudes toward testing and achievement of basic skills, (4) strategies for improving reading comprehension, and (5) school and classrom evaluation. (MM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED233325

Testing Achievement in Basic Skills

Volume I: Executive Summary and Research Reports

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Harry Singer

University of California, Riverside

Robert B. Ruddell

University of California, Berkeley

John McNeil and Meri C. Wittrock

University of California, Los Angeles

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Harry Singer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A final report for a project funded by the California Policy Seminar and Administered by the Institute of Governmental Services (Mr. John Cummins, Director), University of California, Berkeley.

1982

CS 007209

Table of Contents

Page

Foreword

Executive Summary

Testing Achievement in Basic Skills 1

Need for This Study 2

Current Testing Programs 4

Research on Tests and Testing 6

State Testing 6

School District Testing 7

Classroom Testing 9

Recommendations 11

Statewide Testing Program 11

District Level Testing 18

Classroom Level Testing 21

Footnote 26

References 27

Research Reports

State Testing

Test preferences and competencies of field educators 30

Robert B. Ruddell and
Charles K. Kinzer

Knowledge and attitudes toward testing: Field educators and legislators 39

Robert B. Ruddell

Table of Contents
(contd.)

Page

Evaluation and testing of reading: Educators' perceptions, attitudes, and values 51

School District Testing

The great educational earthquake of 1981: Reallocation of resources in California schools to meet minimal competency requirements. Was the reallocation necessary? 61

Harry Singer

Preferences to testing and reporting test results in basic skills: Acquisition vs. learning from text, content-bound vs. content-free 78

Harry Singer and
Mariam Jean Dreher

Attitudes towards testing and test results in achievement of basic skills 112

Harry Singer and
Mariam Jean Dreher

Classroom Testing

Summarization strategies for improving reading comprehension 140

John McNeil and
Lisbeth Donant

School and classroom evaluation 151

Merl C. Wittrock and
Helen J. Schultze

Foreword

Although educators throughout history have used tests of one form or another, only in the last 100 years have standardized, objective tests been developed and employed systematically on a large scale in education. This phase of testing began in reaction to the excesses of the spoils system following the Civil War. Congress voted to establish a civil service which began to devise tests for the objective selection of candidates without regard to race, religion, or social origin. In 1897, a spelling test became the first standardized achievement test. Over the next 20 years enough achievement tests were devised to provide instruments for the scientific movement in education that promised to make schools more efficient by grouping students for more appropriate instruction. Another type of test, one for assessing intelligence, developed by Binet and Simon in 1905-1911 was imported, improved by Terman at Stanford in 1916, named the Stanford-Binet, and used to assess students' capabilities for school achievement.

New York has used its own Regents Examination for over 50 years to assess achievement at the high school level. But California did not begin to use state mandated tests until 1961. Over the next 20 years it employed tests to assess achievement in basic skills, monitor local school districts' curricula, certify competency of students and teachers, and use the information to determine allocation of resources.

Tests have added to the State's control over education. The control already included establishment of local school boards to administer the State's Education Code; textbook selection; credential

requirements for teachers, administrators and specialists; and funds for categorical programs. Most recently, after the passage of Proposition 13, the state began to provide bail-out money for local school districts who could no longer depend upon property taxes as their main source of support. Thus the state legislature has assumed greater control over schools at the cost of local control.

However, this report focuses on only one aspect of legislative control over education: testing achievement in basic skills. But the scope of our inquiry is broader. We have also included information on testing required by the federal government, the school district, and classroom teachers. We also touch on aptitude and achievement tests required for admission to the state's colleges and universities. We do not critically evaluate the college entrance tests, but we point out that a gap exists in information gathered at the high school level for improving the testing program because the state only monitors achievement in basic skills and requires minimal competency testing, but does not assess achievement in college preparatory programs.

Concern over the testing program led the California Policy Seminar to commission a study on "Testing Achievement in Basic Skills." The study consists of two parts: (1) a description and analysis of the current testing program and (2) research on several aspects of achievement testing. A comprehensive report on both aspects, including our recommendations for the improvement of the testing program, is contained in one of our two final reports, Testing Achievement in Basic Skills: Volume II: Description, Analysis, and Recommendations. However, only brief statements of the research results could be included in it. All the research studies have been submitted for

publication in professional journals. In Volume I of our final report on the project, we summarize the research studies, list the recommendations that follow from them, and then present the studies in their complete journal publication form.

In working on this study we have received help from many people. Members of the California Assessment Program, Drs. Alex Law, Dale Carlson, Beth Breneman, and Mark Fetler, were most generous in explaining the CAP program and providing us with tests, materials, reports, and even statistical analyses of CAP data. If we err in this report on any facts pertaining to the CAP program it is not due to any unwillingness on the part of the CAP staff to provide us with the necessary information. We also want to thank the many legislators, their staffs, and the public school personnel and students who responded to our questionnaires, gave us interviews, and participated in our experimental activities for determining test preferences and attitudes towards testing and test reporting. We also received help from the San Diego Unified School District. Dr. Bob Bennett, director of curriculum, provided data which showed that San Diego had expanded its time on task in basic skills and thereby had narrowed its elementary curriculum over the last 5 to 10 years. Dr. Steve Isaacs, San Diego's Director of Research and Evaluation gave a lengthy and detailed response to our request for a school district's view on testing; much of his response is reported in our Final Report, Volume II, Section III. Dr. Robert Hocker, now acting Superintendent of Alvord Unified School District, made arrangements for testing and communicated considerable information about the CAP program and school district testing. Dr. Ross Green of McGraw Hill/California Test Bureau

provided us with his company's tests and other information on tests and testing. We are also grateful to the colleagues, research assistants, and secretaries who worked on aspects of this report. They are Dr. Douglas Mitchell, University of California, Riverside; Dr. Jean Dreher, University of Maryland, College Park; Dr. Charles Kinzer, now at Vanderbilt University; Richard Speaker and Marilyn Draheim, graduate students at the University of California, Berkeley; and Lory Furth and Helen Schultze, graduate students at the University of California, Los Angeles.

Finally we express our gratitude to the California Policy Seminar for awarding us this project and Mr. John Cummins and his staff at the Institute of Governmental Services that oversees the project. We dedicate our report to all the students in California's schools; we hope that the report will help make their education better through improvements in tests and testing practices at the state, school district, and schoolroom levels.

HS, RR, JM, and MW

August, 1982

Executive Summary

Testing Achievement in Basic Skills[a]

The purpose of this report is to evaluate the current testing program, conduct experimental studies on new ways of testing and reporting test information, and make recommendations on how the testing program can be improved in California. As the report shows, we think that the tests, the ways in which the state administers and uses them, the confidence placed in the validity of test information, and how test data are reported should be changed in ways that will make achievement testing in basic skills more beneficial to all those concerned with public education. The evidence for our conclusions and the basis for our recommendations stem from descriptions and analyses of tests currently used in California schools, reports on tests and test results for California students, and seven studies that we conducted for this report.

This summary consists of four parts. The first is an explanation of the need for doing this study, the second is a description of tests in current use, the third contains summaries of our research results, and the fourth is a list of our recommendations.

Need for This Study

Tests not only assess pupils' progress and performance, they also serve for selection and placement of students in programs, and influence students' self-perceptions. Moreover, they reflect upon the competence and efforts of professional educators. Tests operationally define the goals of education, tend to control and regulate what is taught and emphasized in the curriculum, and provide information that influences public policy decisions on education. They also inform the public on the efficiency and effectiveness of the schools in attaining educational goals. Hence tests serve schools and society in many ways.

Testing achievement in basic skills is also an expensive enterprise. California's school districts spend over 40 million dollars each year on assessment. This dollar amount does not include the prorated costs of teacher and student time nor any other indirect costs.

Teachers at best are ambivalent about the testing program. They perceive that the tests use up expensive instructional time. They know that the state-wide testing program cannot provide individual test scores to teachers because each student takes only a sample of test items, not a complete test. Hence the state can only return information on grade levels, schools, districts, and state trends on basic skills. This information cannot be related to individual students. Although the information may be useful to the state, the district, and school principals, it has little meaning for classroom teachers.

Information on the district's standardized tests, which are usually scored at some central location, takes about 3 months to reach teachers. After this time interval, teachers do not have much interest in the test data, particularly in primary grades where student development in achievement is relatively rapid. However, standardized test information from commercially published tests, in contrast to statewide data, informs school boards, parents, principals, and pupils on achievement in basic skills in comparison with national norms. The information that teachers find most instructionally relevant is obtained from the results of their self-constructed or criterion-referenced tests. Students know these tests are related to their grades and the results will be returned to them; therefore, they are likely to be motivated to do their best.

A criterion-referenced test has items or questions related to a particular instructional objective. Ideally students should know the answer to few items before instruction and most items after instruction. The teacher sets a criterion of the percent of items students must get correct to indicate they have learned the objective. Thus, a criterion-referenced test is designed to be instructionally sensitive.

In comparison, a standardized, norm-referenced test also consists of items or questions related to a particular objective(s), but the test-maker, usually a commercial publisher, then gives the test to a sample of students and does an item analysis of the results. Those items are eliminated that all students fail or pass. Standardization consists of retaining only those items that fit certain statistical

criteria. When the test is given to another sample of students, the resulting scores form the comparison norms for the test.

Because of the multitude of tests California students now take, the significant role they play in school and society, the large sums of money expended on them, and the ambivalent attitude teachers have toward tests, they need to be scrutinized objectively and critically.

Current Testing Programs

California students take state-mandated, school district, and teacher-made tests from grades 1-12. The state-mandated test, constructed and administered by the State Department of Education's California Assessment Program (CAP), is called the Survey of Basic Skills (SBS).[3,4,5] It is given annually to each public school student enrolled in grades 1, 3, 6 and 12. School districts must also test students on nationally standardized tests constructed by various publishers. The most widely used standardized tests in California are the California Test Bureau/McGraw Hill's Comprehensive Test of Basic Skills, the Psychological Corporation's Stanford Achievement Test, and the Science Research Associate's Iowa Tests of Educational Development. School districts must administer these kinds of standardized tests to satisfy requirements for students enrolled in federally-funded Title I programs. Some districts also administer standardized tests annually to all its students in each grade. Other districts such as the Los Angeles Unified School District administer

its own test of basic skills to evaluate student progress. Students in all schools also take teacher-made tests which assess achievement of specific instructional objectives. Those high school students who want to apply to the California State University and College system or the University of California must take the Educational Test Service's Scholastic Aptitude Tests. For the University of California, they must also take the College Entrance Examination Board's Achievement Test(s).[1]

Another state-mandated test that all California students must pass to receive a high school diploma is called the proficiency or minimal competency test.[13] It assesses achievement in reading, writing, and arithmetic. According to state law, school districts must select or construct their own proficiency tests, decide the score necessary for passing the test, and administer them once each year in grades 4-6, 7-9, and 10 and 11. Students who pass the tests in lower grades do not need to take the tests again.

The kinds of questions we sought to answer in this report about the testing program are: What use is made of the test results? Do they help legislators decide how to allocate resources? Are they useful to administrators and teachers? Are there other ways of teaching, testing, and reporting test results that would better serve the purposes of education in California? To answer these questions we not only read and evaluated published reports on the testing program, we also conducted seven research studies.

Research on Tests and Testing

State Testing

Interviews of a representative sample of educators, including administrators and teachers throughout the state, revealed that as the interviews progressed from the district to the classroom level, the less was known about the CAP test, how to interpret its results, and the usefulness of its information. Administrators tended to report the test data were somewhat useful, but wanted fewer skills tested; teachers indicated the results were only of marginal use or not useful at all.[7,8,14,15]

Legislators and members of their staff also had an ambivalent attitude toward the CAP program. They reported that while educators are responsible for achievement in basic skills, legislators need achievement information for making decisions on allocation of funds to schools. Most legislators interviewed, however, have little confidence in the objectivity of the state test results. They recognize that the State Department of Education which is responsible for administering and providing leadership for the state's educational program has an elected superintendent whose Department of Evaluation and Assessment is in charge of all aspects of the CAP test from construction of the test to interpretation and release of the results to the public and the press. In effect, the Superintendent is required to evaluate his own program. Legislators realize that this requirement may lead to a conflict of interest. Because of this fact alone, some legislators do not have confidence in the test results and

their interpretation. Several legislators recommended that a contract should be made with an independent group to administer and interpret the assessment program.[14]

School District Testing

Under the Educational Code, school districts must select or construct their own proficiency test. Some 70 percent of the school districts in California elected to construct their own test. Acting under guidelines from the State Department of Education, school districts constructed tests and decided upon a passing score that resulted in some 30 percent or 100,000 10th graders failing one or more proficiency tests in 1981, the first year the test had to be administered under state law. School districts estimated that some 5 percent or 30,000 students will still fail the test when they reach the 12th grade; consequently they will not receive a high school diploma.[13] They will therefore not be eligible for any job that requires one. Since the law requires schools to provide instructional programs for students who fail the test and continue to provide them until the students pass, schools had to reallocate resources to meet this requirement. An estimated $12,000,000 in school funds had to be shifted from the 70 percent of the school population who passed the tests to the 30 percent who failed one or more of them.[16] Since many of the tests involve application of basic skills to life-tasks, such as filling out a rent receipt, reading an advertisement, or computing a grocery bill, that are not usually taught as such in school,

questions have been raised about the legality of the tests.[10] Already some parents have taken school districts to court over due process issues raised by the tests.

School districts also administer standardized tests of basic skills. These tests are well-constructed, provide reliability and validity data, and have national norms. However, these norm-referenced scores are difficult for teachers to use. We conducted an experiment which demonstrated that teachers assign different letter grades to identical student scores that were first reported in percentiles and then in percentages. Teachers seem to use a 100 point scale for evaluating test results as though both types of information indicate percentage of items correct. In other words, they react to the numbers alone, not to what they mean. This experiment indicates that teachers need student scores reported in a familiar and understandable format which will be useful for their diagnostic decisions and instructional plans.[22]

School districts do not report test results in a meaningful way. They usually only indicate percentile rankings from year to year. Lines connecting these rankings are usually flat unless a student has improved more rapidly than the norm group, then the line slants upward. A student may improve but at a slower rate than the norm group, then the line slants downward. These types of reports are susceptible to misinterpretation. In all three cases, the student may have improved, but the public would interpret only the upward slanting line as an improvement. Moreover, districts and schools report scores in numerical form only. Because such reports are content-free, they do not inform the public on what students know and can do. When

percentile rankings are given on the same test at the beginning and end of the year and compared with the same norm group, then a line connecting the resulting percentile rankings is likely to show an upward slant, particularly in the primary grades where basic skills are being acquired at a rapid rate. These rankings should also be shown with content- bound information, for example, the most difficult paragraph a student could read successfully each time the student was tested. If such information were reported, the public, press, parents, and pupils would perceive primary grade students were developing rapidly in acquisition of basic skills. If this self-comparison data were also accompanied with the usual peer-comparison data and with content-bound information, educators would find the information is diagnostically useful and would change their attitudes to testing from negative to positive. In short, educators are not opposed to testing per se, but to the way that tests are currently administered and reported.[17,18]

Classroom Testing

We found from questionnaire data for third and sixth grade teachers that they are ambivalent about tests and test results. A majority indicated tests are important and use test results for making relevant decisions. However, the teachers reported that individual test data are most useful for making classroom decisions.[22] However, teachers can only obtain this information from district

administered standardized tests and teacher-made tests, not from the state administered tests.

Teachers prefer test data that are related to what their instructional materials emphasize and what standardized tests assess. They also want information on the newer processing strategies for attaining reading comprehension, but these strategies have not yet been incorporated into teaching materials and tests. In short, teachers prefer coherence between what they teach and what they test.[22]

Teachers also favor criterion-referenced to norm-referenced data, even when the norm-referenced data would be more appropriate, and tend to perceive students more favorably when their scores are reported in a criterion-referenced format, in percentage of correct items. These observations indicate a need to explain more effectively to teachers the purposes of tests and interpretations of test results, in addition to the need to present test results in a clear and understandable format.[22]

To determine the kind of diagnostic information and the type of reporting format teachers prefer we gave teachers and administrators four types of test information on the same second grade student. They most preferred the diagnostic report that differentiated rate of reading acquisition from peer group comparison and provided content-bound information, that is, the most difficult paragraphs the student could read successfully each time the student was tested. The

educators least preferred the report which listed peer group comparisons only, the type of report that is usually given for standardized testing![17] An attitude scale indicated educators had a favorable attitude toward the diagnostic report, but a negative attitude toward the peer group comparison only type of report.[18]

Because instruction and testing are interrelated, we conducted an experiment to determine what effect teaching one of the newer comprehension processing strategies would have on the state's test of comprehension. We taught elementary students how to summarize what they read. They improved significantly over another group that was asked to summarize what they had read before they were tested, and both groups achieved more on the CAP comprehension items than a group which only read the passage and then answered the questions. In short, more appropriate comprehension instruction resulted in significant improvement on the comprehension test.[11]

Our analyses of current test practices, test reports, and our own experimentation with testing has led us to formulate recommendations for the improvement of achievement testing of basic skills at the state, school district, and schoolroom levels. We precede each recommendation with a brief rationale.

Recommendations

Statewide Testing Programs

Rec. 1. Retain, but modify CAP's Survey of Basic Skills Test for

monitoring state, district, and school performance but include newer types of process oriented skills. Administer CAP to only a small sample of representative schools within school districts each year. Change the title of the CAP report from "Student Achievement" to "School Achievement." Authorize an independent agency, not the State Department of Education, to direct, administer, and interpret the statewide assessment program.

Each administrative level must collect appropriate information for use in making intelligent decisions. The State Department of Education must gather data on how California as a state is performing relative to other states in achievement of basic skills (Education Codes 60600, 60663 and 60640). Although the CAP test meets the criteria for a well-constructed test, it has its strengths[9] and limitations.[2] Its major strength is to provide the state, school districts, and schools with comparative data on achievement in basic skills. Its major limitation is that it does not provide individual scores; even if it did administer individual tests, the Educational Code (60607, 60641) forbids annual reporting of state-determined individual scores or relative positions of individual students to district governing boards; however, the State Board of Education may determine which, if any, of the achievement test results in grades 3, 6, and 12 may be recorded on students' cumulative records (Education Code

60612, 60641). However, the Education Code (60605) permits school districts to administer other tests, including tests that yield individual scores.

The CAP tests also have other limitations. We have already discussed its need to incorporate a more effective way of measuring processing skills. The CAP tests have been recently broadened at the sixth grade to include reports not only on basic skills but also on performance in content areas of science, social studies, and literature.[5] This broadening is likely to counteract a tendency for elementary schools to narrow their curricula to emphasize only basic skills. This emphasis may have been counterproductive to achievement in basic skills, particularly to reading comprehension.[12] A similar broadening of the CAP tests should be done when the 12th grade tests are revised.

The CAP test also lacks national norms. Unfortunately, there are no test norms on a state-by-state basis. The only way such norms could be constructed is to administer the CAP tests to a representative sample of schools in each state. We do not foresee such norms forthcoming. To approximate a comparison with other states, California equates its test to nationally standardized test results, but caution must be observed in interpreting these results because the theoretical narrowness of state-by-state norms would magnify greatly any state's departure of test results from the average on the hypothetical state-by-state norms.[19]

To collect information on state trends, it is not necessary to test every student in grades 1, 3, 6, and 12. The state could obtain the same result from a sample of schools in each district.[21] This sampling approach would be more economical than the current practice. Each school would then not be tested on the state test, but almost every school already tends to get similar data from standardized tests administered by local school districts.

The state could continue to collect and equate district standardized test results with the state test. The state currently uses such standardized test data to equate the CAP test with national norms. Most school districts are now using standardized tests to satisfy federal requirements for Title 1 programs.[19]

Our recommendations should be viewed in an historical context. In 1961 the legislature passed a law requiring the State Department of Education to provide data on achievement in California schools. The state implemented the law by collecting school districts' standardized test results. But difficulties in equating results from diverse tests led the state in 1962 to adopt a single standardized test for its statewide testing program. After some seven years of successive commercially-published test adoptions and use, the legislature responded to district objections over the validity of the tests by passage of the California School Testing Act of 1969 and by further legislation in 1972 that, among other changes, allowed the State Board of Education to construct its own test. The scores of the testing program were to be "limited primarily to the task of furnishing information useful to state policymakers and decision makers;"

it was assumed that "assessment information needed at the classroom level could best be collected by local school personnel. In spite of this assumption, the [CAP] program was designed to report as much information as possible to local personnel" (p. 6).[4] Our recommendation reinforces the legislature's limitation on the scope of the CAP program and advocates that the legislature's assumption be fulfilled. The CAP program would then only serve its intended purpose and would avoid problems that arise when it attempts to make its results useful to local personnel.

Although the state solved some problems in making the 1972 changes, such as increased test breadth and relevancy and reduced testing time, the change created other problems, such as having a test that does not provide individual scores and overlaps with school districts' standardized test results. Also, since the state continues to interpret its tests through the use of commercially published standardized test norms, it has not solved one of its original complaints that commercial tests have out-of-date norms. Our recommendation reduces the overlap, makes the state testing program more cost-effective, and solves the initial problem of how to compare results from diverse standardized tests. The state can now equate the standardized tests used in local school districts to the common CAP test.

The CAP tests should also be modified in another important way: they should be referred to as "School Achievement," not as "Student Achievement" data. Since no student takes a complete test and since no score can be reported on individual students, the term "student achievement" is misleading.[19]

More important is the lack of confidence of legislators in the CAP test results.[14] To spend millions of dollars on a testing program and not to have confidence in the results and their interpretation because of a possible inherent conflict of interest is cause enough for changing the CAP program. We therefore recommend that legislators modify the Educational Code in such a way that an independent agency can be contracted to administer the CAP program and satisfy the objectives of providing information to the legislature that is necessary for making decisions on allocation of resources, evaluation of programs, and identification of unusually successful practices and factors which appear to be responsible for attainment of high achievement (Education Code 60601).

Rec. 2. Devote resources of the State Department of Education to construction or selection of and experimentation with new tests and related materials and instructional procedures that may be useful for improving achievement in basic skills.

The State Department of Education has the resources and the capability to serve this function, which no one else is doing. For example, knowledge and procedures for assessing process-oriented skills, such as summarization have recently become available and have improved students' reading comprehension.[11] Assessment of this kind of skill will provide teachers with needed information on whether students are acquiring necessary ways of attaining achievement in basic skills.

Test publishers already know that process-oriented skills can be taught, learned, and assessed. But they are reluctant to incorporate the new skills in their test battery lest schools find the change too great and not purchase the test battery. Consequently, they only tend to assess current curricular and instructional objectives, as manifest in materials. Thus, they are assured their tests will have curricular and instructional validity. Teachers tend to teach what is in their instructional materials and want tests that assess what they have taught. To break this circular process, the State Department of Education will have to experiment with new materials that develop the new processes involved in attainment of achievement and use new tests to assess whether students are attaining the process. Apparently both tests and instructional materials have to change simultaneously for teachers to change what they teach and what they test or want to have tested.[22] Indeed, learning, instruction, and testing should be coherent.[20]

When the new tests are shown to be valid, they can be incorporated into the California Assessment Program's test battery. Thus, through curriculum innovation and assessment, the State Department of Education will be contributing to its leadership role in education.

Rec. 3. The state legislature should gather information on the level and quality of performance of those students enrolled in college preparatory courses required for admission to the University of California, California State University and Colleges, and to other public and private universities.

While analyzing the existing testing program, we became aware that information on school achievement that the state now obtains is inadequate for making decisions about allocation of resources at the high school level because the state does not monitor its mandated programs for college preparation. Apparently college preparatory standards in California are lower than the national average. For example, the College Entrance Examination Board reported that California students take fewer academic courses for less time and get higher grades than students in general throughout the nation.[1] However, if the legislature obtained achievement data on the college preparatory program by administering its own tests or relying on achievement test results from the College Entrance Examination Board tests, it could use such information for determining allocation of resources. It would then have information on all three tiers of high school students: those in the college preparatory program, those who still need to pass a minimum competency test, and those students in-between who represent the average performance in basic skills as assessed by CAP tests.

District Level Testing

Rec. 4. School districts should not only assess achievement in basic skills, but also in the content areas of science, social studies, and literature.

If school districts only assess achievement in the basic skills, teachers are likely to focus on development of these skills and thus narrow the curriculum. The result is that students would not attain as much general knowledge. This knowledge is necessary for reading comprehension. Most printed materials require students to draw upon general knowledge for making inferences as they read and for interpreting printed materials. The consequences of a narrow curriculum are more likely to be observed in the higher grades. The remedy is simple: assess students in both basic skills and in the content areas of science, social studies, and literature.

If standardized tests are used for this assessment, the results will also provide boards, parents, professionals, pupils, and the public with information on the relative standing of students, classes, schools, and districts in relation to national norms. We know that teachers need this information to help them make informed and objective decisions in selection and placement of students.[22] Fortunately, most districts are already providing this type of information.

Rec. 5. <u>Districts should retest students in the primary grades on the same tests and use the same norms to interpret the scores.</u>

Self-comparison data that results from retesting is likely to demonstrate that primary grade students generally make rapid progress

in acquisition of basic skills. This kind of information will inform legislators and the public on the degree of success that schools are experiencing in developing students' acquisition of basic skills.[17]

Rec. 6. Report all test results with sample tasks.

The sample tasks will enable the test consumer to see what students know and can do on each type of test at each level of testing.[17,18]

Rec. 7. The legislature should allocate funds for high schools to carry out the mandated remedial program for minimal competency attainment.

Given the authority, but with guidance from the State Department of Education, to construct or select their own minimal competency tests, school districts tended to set the standard at the ninth grade level with 70 percent as the passing score. Some 30 percent of the ninth grade students who fail the tests have to take one or more remedial classes at high school. School districts have had to reallocate their own resources for the mandated program since the state provides no additional funds for it.

However, some school districts set the standard for their minimal competency test at the sixth grade level where they were prior to minimal competency testing. They established a passing score at such

a level that at least 70 percent of the students in the sixth grade could pass the tests. Almost all students could pass these tests prior to high school. Consequently, their high schools had very few remedial courses in basic skills and almost all the students who completed 12th grade passed the test and received a high school diploma. Thus, these high schools did not have to reallocate much resources and do not need to seek much, if any, reimbursement funds.

If school districts want to adopt a sixth instead of a ninth grade standard, they should know that they already have the authority to make this change in their minimal competency tests.[16] These districts should also inform their boards that minimal competency requirements should not be confused with graduation requirements.

Rec. 8. The legislature should require each school district to adopt procedures for improving its administrators' and teachers' ability to (1) construct, administer, and interpret criterion-referenced and norm-referenced tests and (2) use test results for making decisions on curriculum and instruction. The State Department of Education should be authorized to provide school districts with guidelines on how to effect this improvement. These guidelines should be updated as new information on tests and testing becomes available. Each school district should then report to the State Department of Education how it is implementing this recommendation.

Our research has indicated that school district administrators and teachers need to have more knowledge on all aspects of testing[8,14,15] With more adequate knowledge, they will be better qualified to adopt or construct tests, administer them to students, interpret the results, and apply the knowledge gleaned from the tests to decisions-making processes in curriculum and instruction. Adequate knowledge of tests and testing should also enable them to beware of the pitfalls of assessment, particularly on test bias, in relation to minority groups.

We think that each school district should have the responsibility for deciding how it will implement this recommendation under guidelines from the State Department of Education. In providing guidelines the State Department should state the various ways in which this recommendation can be implemented and let each school district select an option or devise an alternative that will produce the same result. School districts should be encouraged to work with personnel in local colleges and universities in designing this plan. Each school district should then inform the State Department of Education of its decision and describe how it is implementing this recommendation.

School districts must also keep up with change in tests and testing as they occur. For example, new tests are likely to become available on assessment of knowledge a student has acquired, use of this knowledge, and diagnosis of what has been learned and what still needs to be improved in subject matter learning.[6] By incorporating

such information into revisions of its guidelines every few years, the State Department of Education will be helping local school personnel keep up-to-date on achievement testing in basic skills.

Classroom Level Testing

Rec. 9. Provide teachers with test data on individual students, both criterion and norm-referenced. These data should be diagnostic and reported in a meaningful way so that they can be used for instructional decisions at the classroom level.[16,17,22]

Teachers need criterion-referenced data for making instructional decisions. They also should have norm-referenced data for selection and placement of students.

If students, particularly in the primary grades, are retested on the same tests of basic skills after a year of instruction, as well as on a norm-referenced test standardized for the grade level, then diagnostic information would be available on each student's progress in acquisition of the skill and relative comparisons with a peer group.[16]

To make reports meaningful, they should contain information on percent of items correct and percentile rank. These two types of information are complementary. Reports for each student should also contain examples of the most difficult task the student successfully completed on each basic skill.[17,18]

The tests, both criterion and norm-referenced, should assess not only what students know but also how they acquire this knowledge. Hence, tests need to contain processing-oriented as well as knowledge scores.[11,22]

Elsewhere we have recommended that the resources of the State Department of Education should be devoted to construction of and experimentation with new types of tests and with preparation, selection, and try-out of new materials and instructional procedures to go with experimentation in new types of tests. To effect change in teaching of basic skills, it is necessary to <u>modify both tests and materials simultaneously</u>.

However, we would caution schools not to proliferate the number of criterion-referenced tests administered to students. Such proliferation could result from defining instructional domains narrowly and assessing each domain. For example, defining a domain of types of commas in punctuation, instead of a more general domain of punctuation. Only those domains should be assessed in which teachers need and will use information for making instructional decisions. All those involved with authority over testing students, from the state to the classroom level, need to exercise wisdom and judgment in exercising this authority.

With our recommendations in force, we believe that the state will have more adequate, more dependable, and more trustworthy data for making decisions on allocation of resources, and will be providing itself, parents, pupils, press, and the public with more meaningful reports on achievement in California's public school.

Footnotes

a. The second volume is called Testing Achievement in Basic Skills: Volume II: Description, Analysis, and Recommendations. A final report to the California Policy Seminar, Institute of Governmental Services (Mr. John Cummins, Director), University of California, Berkeley, California, 1982.

References

1. Baker, W. and A. Cox (Co-Chair). Report to the Policy Committee on the University of California's Activities to Assist Underprepared Students. Part I: Descriptions of Trends and 1979-80 Activities; Part II: A Compendium of Ideas for Further University Activity. Berkeley, California: University of California, March 1981.

2. Balow, I. H. But, does CAP help school districts? California School Boards, September 1978, 37(7), 18-21, 34.

3. California Assessment Program. Technical Report. Sacramento, California: California State Department of Education, 1977.

4. California Assessment Program. Student Achievement in California Schools: 1979-80 Annual Report. Sacramento, California: California State Department of Education, 1980.

5. California Assessment Program. Survey of Basic Skills. Separate reports on Rationale and Content for Grades 1, 3, 6, and 12. Full Report: Part I, Survey Results; II, Program Diagnostic Displays; III, Interpreting and Using Test Results; IV, Skill Area Descriptions. Also separate Examiner's Manual for each test level. Sacramento, California: State Department of Education, 1980, 1982.

6. Glazer, R. The future of educational research. The Educational Researcher, 1982, 11(8), 16-17.

7. Kinzer, C. Field educators perceptions, attitudes, and values on testing reading: An international perspective. Paper prepared for the National Reading Conference's Annual Meeting, St. Petersburg, Florida, December 1982.

8. Kinzer, C. and R. B. Ruddell. Evaluations and testing of reading: Field educators' perceptions, attitudes, and values. Paper presented at the National Reading Conference's Annual Meeting, Dallas, Texas, December 1981.

9. Law, Alexander I. CAP - Statewide testing comes of age. California School Boards, September 1978, 37(7), 6-22.

10. McClung, M. S. Legal implications of intelligence and competency testing. Center for Law and Education, Cambridge, Mass.: Harvard University, 1979.

11. McNeil, J. and L. Donant. Summarization strategy for improving reading comprehension. In J. Niles and L. Miller (Editors), New Inquiries in Reading Research and Instruction. (Thirty First Yearbook of the National Reading Conference.) Rochester, New York: The National Reading Conference, 1982.

12. McNeil, J., H. Singer, and L. Furse. Curricular scope as a causal factor in reading achievement. Paper prepared for the National Reading Conference's Annual Meeting, St. Petersburg, Florida, December 1982.

13. Proficiency Assessment in California: A Status Report. Sacramento, CA.: State Department of Education, 1980.

14. Ruddell, R. B. Knowledge and attitudes toward testing: Field educators and legislators. Paper prepared for the National Reading Conference's Annual Meeting, St. Petersburg, Florida, December 1982.

15. Ruddell, R. B. and C. K. Kinzer. Test preferences and competencies of field educators. In J. Niles and L. Miller (Editors), New Inquiries in Reading Research and Instruction. (Thirty First Yearbook of the National Reading Conference.) Rochester, New York: The National Reading Conference, 1982.

16. Singer, H. The great educational earthquake of 1981: Reallocation of resources in California high schools to meet minimal competency requirements. Was the reallocation necessary? The California Reader, March/April 1982, 15(3), 13-19.

17. Singer, H. and M. J. Dreher. Preferences for testing and reporting test results in basic skills: Acquisition vs. learning from text, content-bound and content-free. In J. Niles and L. Miller (Editors), New Inquiries in Reading Research and Instruction. (Thirty First Yearbook of the National Reading Conference.) Rochester, New York: The National Reading Conference, 1982.

18. Singer, H. and M. J. Dreher. Attitudes towards testing and test results in achievement of basic skills. Final report, submitted to California Policy Seminar Project: Achievement Testing in Basic Skills. Berkeley, California: Institute of Governmental Services, 1982.

19. Singer, H., R. B. Ruddell, J. McNeil, and M. C. Wittrock. Testing Achievement in Basic Skills: Volume II: Description, Analysis, and Recommendations. (Final Report to California Policy Seminar.) Berkeley, California: Institute of Governmental Services, 1982.

20. Tyler, R. W., & S. H. White. Testing, Teaching and Learning. (Report of a Conference on Research on Testing, August 17-26, 1979) Washington, D. C.: National Institute of Education, 1979.

21. Utilization of the California Assessment Program by Local School Districts. Los Angeles, California: Southern California School Districts Superintendents' Council, November 1978.

22. Wittrock, M. and H. Schultze. School and classroom evaluation. Final Report on California Seminar Project, Testing Achievement in Basic Skills. Berkeley, California: Institute of Governmental Services, 1982.

RESEARCH REPORTS

State Testing

Test Preferences and Competencies of Field Educators

Robert B. Puddell

University of California, Berkeley

Charles K. Kinzer

Peabody College of Vanderbilt University

Testing plays a major role in interpreting academic achievement of schools to the public and state legislatures throughout the nation. Little information is available, however, to indicate how the classroom teacher, principal, and district office personnel use test data, their technical knowledge of testing which affects interpretation of data, and their perception of the effect of testing on the curriculum in schools (Singer, et al., 1980). These areas were of central importance in this research. The following research questions served to guide the study*:

1. How are tests used by teachers, principals and central office personnel?
2. What technical test/knowledge is possessed by school personnel?
3. What impact does testing have on the school curriculum at the school site and district wide level?

This research was conducted in the state of California, where like other states, schools utilize tests developed by the State Department

*This research was supported by the California Policy Seminar, University of California (John Cummins, Director).

of Education. The intent of the California Assessment Program (CAP) is to provide information to the legislature and the public on basic skills achievement of California students. The program is designed to compare California students with student achievement nation-wide, to assist in decision making in distribution of funds, and to attempt to identify causal factors which contribute to basic skills achievement. To conserve testing time, a multiple matrix sampling procedure is used. With the exception of the grade one entry test no students responds to more than ten percent of the items on any one test, requiring that a composite group test score be derived for a given district. No classroom or individual scores are reported and within the district the minimal unit for reporting is the individual school site. In addition, however, most schools administer a standardized achievement test selected at the district level in order to provide individual and classroom evaluation data.

Method

The population for the study consisted of eight school districts selected on the basis of size, location, and student composition. Two districts represented large urban populations with high percentages of minority students identified as low SES based on data provided by the State Department of Education. Two districts were of median size in a suburban setting representing a middle SES level. One district was of small size, suburban, and of high SES student composition. Finally, one district was of small size, in a rural location at the mid SES

level. These districts ranged in location from Southern California to the San Francisco Bay Area and rural Northern California.

Following the formulation of the research questions a 25 item questionnaire and a detailed interview schedule were constructed, field tested, and revised. Combined time for the administration of both instruments consisted of approximately one hour.

Arrangements were made with the central district office in each district for interviews with teachers at the third grade and sixth grade levels, the school principal in each teacher's building, and central office personnel. District office personnel were asked to establish the interviews with school personnel in a school or schools best representing a composite makeup of the district. Questionnaires and interviews were conducted with thirty eight individuals consisting of 18 teachers, 12 principals, and 8 district office personnel.

Variation in response to test use for personnel within a district exceeded that between districts. An inspection of the data revealed that for analysis and interpretation purposes the data from the eight districts could be pooled for discussion purposes.

Findings

The first research question on test utilization produced results which indicated that the most frequent use of tests was for diagnosis, placement, and measuring student growth. These responses were found with greatest frequency for the third and sixth grade teachers, ranging from 33% to 100% and followed by principals and district office

personnel, ranging from 17% to 33%. Special program placement for the elementary teacher also revealed a higher frequency, 22% to 44%, than for principals and district office personnel, 0% to 17%. In contrast, principals and district office personnel used tests more frequently to determine trends 25% and 50%, respectively, than did teachers, 0%.

Information on use of tests to assess instructional programs revealed that teachers and principals make greatest use of tests in "diagnosis" of classroom programs, ranging from 50% to 87%. Principals expressed greatest concern that a close relationship be present between test items and the curriculum, reporting at a 50% level. The use of tests for program evaluation, not pupil assessment, is of greatest concern to district office personnel, 33%, and to a minor degree the principal, 8%. Both district office personnel and principals express the view that test use is important in providing information to the State Department of Education and the legislature, each at the 17% level.

All school personnel were concerned about specific limitations of tests. The third and sixth grade teachers believed that tests should measure the actual performance of pupils as the reading process occurs, rather than attempt to simulate this process in a test format, 22% and 44%, respectively. Teachers were also concerned that testing not be the sole criterion for measuring progress, ranging from 33% at the third grade to 22% at the sixth grade level. Principals expressed the need for greater availability of subtest interpretation to assist in curriculum planning, 50%, while district office personnel

were most concerned that testing consumed too much time because of multiple test administration, 76%.

The second research question dealt with test knowledge possessed by school personnel. All individuals expressed the view that achievement test performance was closely related to reading ability. Test making skill was identified as contributing to high performance by principals, 33%, and 6th grade teachers, 11%. Most consistent agreement on a single factor affecting high test performance was student attitude toward test taking, with strongest opinions held by third grade teachers, 56%, and principals, 42%. A smaller percentage of teachers believe that test performance is influenced by careful work, 22%, self motivation and confidence, 22%, and parents that value education, 11%. A comparatively high percentage of principals expressed the view that parents who value education influence test results, 42%.

The most important criterion for test selection identified by all personnel was the fit between the test and the curriculum, 58% to 89%. The readability of test items was considered of some importance to third and sixth grade teachers, 22% to 33% respectively, and to principals, 25%. District office personnel expressed greatest concern that the test be easy to administer, 50% and that recent norms be present, 33%.

Test preference responses indicated that teachers favor criterion referenced tests over standardized tests by a 78% margin. The third and sixth grade teachers believe that criterion referenced tests provide a more complete measure of individual growth, 78% to 44%, and

again value items requiring performance which closely parallels the reading process present in their instruction, 67% to 44%, respectively.

While the above information provides some clue to test knowledge the most direct indication of technical test knowledge came from one interview item on the interview schedule. Specifically, a brief section of the California Assessment Program informational report used for local school interpretation was provided each individual and interpretation of the standard error of measurement was requested. Of the 38 individuals interviewed correct interpretation was provided by two teachers, one principal, and two district office personnel. Principals and district office personnel expressed the view that teachers and administrators do not understand technical aspects of testing, 42% and 83% respectively.

The third research question explored the effect of testing on the curriculum. While the majority of school personnel expressed the view point that testing has some impact on the curriculum, 33% of the third grade teachers, 89% of the sixth grade teachers, 25% of the principals and 33% of the district office personnel believe that testing has no direct impact on their curriculum. Teachers expressed the view that tests are used to reaffirm teacher evaluation, 33%, with principals at the 17% level. The belief that test results were so general as to provide no help was reported by 33% of the third grade teachers and 44% of the sixth grade teachers.

Specific concern was given to emphasis placed on instructional preparation for achievement testing. A remarkably high percentage of teachers, 33% to 67%, indicated that some form of test taking behavior was taught. This included practicing test taking strategies, 56% to

67%; direct teaching of skills and strategies expected, 44% to 56%; and review of expected questions and formats, 33% to 56%. Conversely, 22% of the third and sixth grade teachers reported that no direct teaching to tests was present. Principals responded at a 33% level and district office personnel at a 67% level that no direct teaching to tests was present.

Conclusions

Based on the findings of the study three conclusions may be drawn. First, in regard to test use the majority of teachers, principals, and district personnel do believe that testing contributes to their curriculum efforts. Teachers and principals use test data which yield results most closely associated with the individual student for diagnosis, placement, and measurement of growth. District office personnel, and to some degree principals, use test data which serve to identify achievement trends for the school district and State Department of Education.

Second, in regard to test knowledge there is a general absence of technical knowledge present for the teacher, principal, and curriculum personnel in the central office. There is, however, an awareness at all levels of key factors which are essential to a good test. These range from a good fit between test and curriculum to the importance of test taking strategies and from the importance of recent norming to readability of test items. Teachers overwhelmingly favor criterion referenced tests in contrast to achievement tests

because data can be more effectively used in identifying instructional problems with the student. Further, teachers believe that test items should approximate normal reading material and test performance requirements should be as close to normal reading behavior as is possible. Both third and sixth grade teachers express the view that tests should not be the sole criterion for evaluation. Principals state that there is greater need for subtest information to assist interpretation, while district office personnel express concern that testing time is too great and moves teachers and students away from direct instruction. The test sense of school personnel appears to have been derived from classroom and test administration experience rather than from any technical study of testing and evaluation.

Third, the majority of teachers and principals believe that tests have impact on the curriculum. There is strong accountability pressure placed on teachers and principals as well as central office personnel. Perhaps then, it is not surprising to find the high percentage of teachers and principals who feel compelled to teach test taking strategies and skills, and in some cases review expected test questions and test formats. One can only speculate on the impact of such instruction on achievement test results. Perhaps the best advice is found in developing more modest expectations for testing and in confronting the realization that achievement testing provides only one data source for instructional decision making.

References

Singer, H., Ruddell, R., McNeil, J., & Wittrock, M. Testing Achievement in Basic Skills. California Policy Seminar, University of California, Berkeley, 1980.

Knowledge and Attitudes Towards Testing: Field Educators and Legislators

Robert B. Ruddell
University of California, Berkeley

Educational funding, like all other forms of public funding, has entered an era of restrained budgeting which has caused the educational establishment to examine all aspects of educational expenditure from the local level to the state and federal levels. In California, with education costing the state between eight and nine billion dollars annually, it is estimated that over forty million dollars is spent annually on testing (Singer et al., 1980, p. 1), a figure that represents one-half of one percent of the annual budget (Analysis of the Budget Bill of the State of California, 1982). This report examines what educators: teachers, principals and other administrators; and legislators know about testing and what attitudes they hold towards testing and test use.

For testing to be effective, it must serve many masters. Thus, the schools have developed testing programs to provide the different types of information needed by the various test users concerned with public education. It further implies that those who are using tests and making policy decisions based on tests be intelligent consumers of the data they receive; the user of the testing instrument must have developed a knowledge of specific types of tests and different types of scores along with a general sophistication in measurement concepts to interpret test results.

Three research questions were formulated to guide this investigation:

1. What test knowledge do educators and legislators possess?
2. What test attitudes are held by educators and legislators?
3. How do testing knowledge and attitudes affect instructional decision making?

II. Design of Study

As part of a larger study (Ruddell & Kinzer, 1981; Kinzer & Ruddell, 1981), parallel interviews were constructed for the purpose of collection data from field educators (i.e., teachers, principals, district office personnel) and legislators. In addition, a questionnaire was developed to be used with field educators. The instruments were piloted and revised. Approximately one hour was required for their administration.

The sample population of field educators was drawn from eight school districts. These districts were selected to represent the broad spectrum of the districts in the state and ranged in location from Southern California to the San Francisco Bay Area and rural Northern California. Two districts were composed of large urban populations with high percentages of minority students at low socio-economic status (SES) level, based on the California Assessment Program (CAP); two districts were suburban of medium size at the mid SES level; one district was rural at middle SES level; and one district was suburban at high SES level. The 38 field educators interviewed from these districts consisted of nine third grade teachers, nine sixth

grade teachers, twelve principals and eight district office personnel.

Selection of legislators was based on two criteria: first, representation from the sampled school districts, and second, representation from key educational legislative committees.

Interviews were conducted with five legislators or their aides, the chief legislative analyst for education and one former legislator who had chaired the Senate Education Committee. This sample included both Democrats and Republicans and three minority individuals. Two of the individuals interviewed were members of the Assembly Education Committee while one had served on the Senate Education Committee. In addition, the top officer from the Office of Program Evaluation and Research in the State Department of Education was interviewed.

III. Findings

The discussion which follows will first examine findings related to formal and informal test knowledge of educators and legislators and their ability to apply test information in instructional decision making. Second, attitudes toward testing of these two groups will be examined.

A. Knowledge

The stimulus questions used to ascertain the formal test knowledge of S's required the interpretation of the scaled scores presented in Figure 1.

The following is an example of the way the state displays the California Assessment results.

SKILL AREAS FOR READING	SCALED SCORE & STANDARD ERROR 1978-79	SCALED SCORE & STANDARD ERROR 1979-80
III. Comprehension	227 ± 07	225 ± 07
A. Literal		
1. Details	223 ± 10	223 ± 10
a) single sentences	233 ± 14	239 ± 14
b) 2 or 3 sentences	240 ± 20	232 ± 20
2. Pronoun reference	218 ± 18	225 ± 18
3. Sequence	208 ± 20	199 ± 20

Figure 1. Test Scores (California Assessment Program, 1980)

In general, the ability to interpret these scores was low. Only 11% of teachers and 17% of principals demonstrated an understanding of the concept of standard error of measurement as shown in Table 1.

Table 1

Understanding of Standard Error of Measurement (SEM)

	Teachers	Principals	District Office	Legislators
Demonstrated Understanding of SEM	11%	17%	37%	17%
Understanding of SEM not Demonstrated	89%	83%	63%	83%

District office personnel demonstrated an understanding of this measurement concept necessary to the interpretation of many standardized tests (e.g., CAP) more often than any other group interviewed but still 63% of these S's did not exhibit an understanding when asked to interpret scaled scores. One legislator with an engineering background correctly interpreted the scores but most (83%) were unfamiliar with the concept.

Decisions based on these test scores varied most with the role of the interpreter. Classroom teachers tended to be most concerned with the drops in the comprehension score and one mentioned that her school board would be unhappy with the change. Principals (50%) tended to mention that they would need to know more about the fit of the test to the curriculum while district office personnel most often (50%) looked for trends.

To examine informal knowledge about tests, S's were asked if they or their peers would have difficulty understanding the test scores and to rank the following scores for utility: total test score, raw scores for each subtest on a larger test, percentile score for each subtest on a larger test, percent correct for each subtest on a larger test, standard score for each subtest on a larger test and grade score for each subtest on a larger test.

Most teachers (89%) indicated that they had no understanding or some difficulty understanding the test scores and would need some assistance in interpreting them. Five principals (42%) indicated that teachers and administrators would have no difficulty understanding the results but since only 17% of principals demonstrated that they

understood the concept of standard error of measurement, it is difficult to have confidence in their estimation of their abilities. In clear contradiction to these principals' opinion that teachers and other principals understand test scores, no district staff indicated that teachers and administrators would understand the scores. Most emphasized the need for interpretation assistance from someone familiar with measurement concepts. Legislators agreed that most of their colleagues would not understand test scores and in fact had little interest in using them. The state department official indicated a particular concern for the legislator's use of test scores and emphasized the need for training legislators in understanding measurement concepts and score interpretation. This official suggested that this could be handled by inservicing freshman legislators at the beginning of their term of office. All groups were asked to discuss the uses they made of test data and the effect tests had on the decisions they made. Teachers made the greatest use of tests in diagnosing (83%) and planning instruction for their students (39%). The most common response of field educators (47%) was that standardized tests had little impact on decisions; teachers were most likely to make this assertion (61%). Teachers (67%) and principals (50%) used tests to diagnose the school's program as did some district office personnel (17%). One legislator indicated that the statewide tests (CAP) should be used by administrators to evaluate and improve instruction but that he felt that this was not being done.

B. Attitudes

To examine attitudes, the S's were asked to discuss their general testing philosophy and to consider what decisions they would make based on the test scores in Figure 1. S's were also asked to rank different tests for different purposes.

Attitudes appeared to change with the S's distance from the child in the classroom. 83% of teachers wanted tests that could be used to diagnose a student's progress; 56% wanted specifically to measure growth rather than just weaknesses. Principals and district office staff mentioned these less frequently. One legislator expressed the need for more effective diagnosis of individual students. Of high concern for the principals was the match-mismatch of the test with the curriculum (50%); personnel at the district and state level were more concerned with trends and the heavy mandating of tests by the state.

Teachers and one legislator stated that tests should not be the sole criterion for placement decisions regarding grade, instructional level or special education.

In general standardized tests were rated as somewhat important but used in the ideal situation could be of great value. All district staff and principals (100%) indicated that in the ideal situation tests would be somewhat to extremely useful while only 84% of teachers expressed this viewpoint.

A theme that ran through the legislative interviews was that the public in general and legislators in particular were not interested

in test scores. Only one individual indicated that test scores had been used in a campaign and that the reports furnished by the State Department of Education had been regularly used to follow the progress of the schools in his district.

Skepticism of testing was indicated in different ways in the legislative interviews. One legislator with a teaching background felt that a testing program should be used to measure the attainment of educational goals and objectives that are realistic in making a person successful but preferred to use tests for diagnosis and prescription that would become the basis for the student's instruction. Another was in favor of using multiple tests for assessing achievement; he felt that the uniqueness of each student required more than one achievement measure. One individual, although skeptical about tests in general, due to the problems of reliability and validity, felt that there was still a need for the development of a fair barometer of student achievement that kept the student in mind. A fourth gave the most pessimistic view of all, he indicated that he thought that the scores published by the State Department of Education were adjusted to serve the purposes of the State Department and could not be trusted because they made the schools look better than they actually were.

Several individuals expressed concern for the low achievement of minority students but were unwilling to go so far as to accept the idea that there should be different tests for different groups of children. In commenting on the different tests issue, one legislator discussed her view that test scores fostered "misguidance" by perpetuating a system of stereotyping that "its the system against the pupil."

She felt that less emphasis on testing would be better for school children if replaced by an instructional emphasis on enthusiasm and motivation.

Different types of tests for different purposes were considered appropriate by all. When ranking tests for statewide use, all groups showed a preference for either standardized or published criterion-referenced tests (CRT's). On a schoolwide basis, teachers did not show a clear preference while principals looked on teacher-made tests with considerably less favor than standardized tests; district level staff preferred CRT's to standardized tests. For classroom testing, no clear pattern emerged in the rankings of any group.

Most S's expressed a strong opinion on the extensive testing required in schools. Teachers commented repeatedly that tests were not really disliked, but that there was simply too much testing of a repetitive nature. The districts in the sample administered the CAP test and at least one other standardized test; larger districts additionally used a district-developed measure. When coupled with everyday testing performed as part of any instructional program, many respondents felt that there was too much time being spent on testing.

IV. Conclusions

Across groups there appeared to be little understanding of measurement terms, although this was more true of teachers than principals or district office staff; legislators indicated that the average legislator would have similar difficulties with measurement concepts.

Scaled scores and standard error of measurement were generally misunderstood.

Those interviewed felt that there was too much duplication of testing in the schools. Principals, especially, felt tests should be reviewed and a single, standardized test chosen, constructed or modified in order to eliminate what they viewed as too much time spent on testing. This is in direct conflict with the view of one legislator who pointed out that a variety of measures must be used to measure the achievement of the students in the California schools because each student is unique. Another legislative respondent expressed such lack of confidence in scores that elimination of testing was suggested. Given the number of tests used in the classroom, the need for reduced testing seems well founded.

V. Recommendations

From this study, three major recommendations can be made. First, due to the general lack of understanding of the formal aspects of score interpretation among field educators and legislators, a program of inservice education must be developed that would train teachers, principals, district office personnel and legislators to understand measurement concepts and interpret test data. This is needed if field educators and legislators are to be intelligent consumers of test data leading to improved instructional decision making and resource allocation. Furthermore, effective communication with the public on the progress of the schools can be enhanced if

educators and legislators are intelligent consumers of test results.

Second, school district personnel should carefully consider the degree of test overlap present in the classroom. Testing time should be markedly reduced with prime concern given to test information which will be useful for instructional decision making by teachers, principals and district office personnel.

Third, changing the responsibility for interpreting and reporting statewide scores from the State Department of Education to an independent agency would help restore legislative confidence in statewide achievement testing.

References

Analysis of the budget bill of the State of California, Report of the Legislative Analyst to the Joint Legislative Budget Committee, Fiscal year July 1, 1982, to June 30, 1983.

California Assessment Program, Survey of basic skills: grade 3 interpreting and using results, Part III, 1980.

Kinzer, C. K., & Ruddell, R. B. Evaluation and testing of reading: Field educators' perceptions, attitudes and values. Peabody College of Vanderbilt University, Nashville, Tenn., 1981.

Ruddell, R. B., & Kinzer, C. K. Test preferences and competencies of field educators. University of California, Berkeley, 1981.

Singer, H., Ruddell, R., McNeil, J., & Wittrock, M. Testing achievement in basic skills, California Policy Seminar, University of California, Berkeley, 1980.

Evaluation and Testing of Reading: Field Educators' Perceptions, Attitudes, and Values

Charles K. Kinzer

Peabody College of Vanderbilt University

Robert B. Ruddell

University of California, Berkeley

Introduction

Even before the current "back to basics" movement, testing and test use has been a focal issue in educational settings. In addition to traditional standardized measures, a proliferation of informal, criterion referenced, district-made and local school assessment measures are commonly used. With so many measures being available, and with the seeming ever increasing emphasis on test scores, it becomes important to find how and why tests and test scores are being employed.

"How" tests are used has been recently reported elsewhere (Ruddell & Kinzer, 1981). This brief paper reports preliminary results dealing with "why" of test use across three groups: teachers, principals, and district office staff.

Sample and Method

Eight school districts in California, chosen to represent small and large enrollments, high and low SES, rural and urban settings, and

inner city and suburban student population, were included in our sample. Data were collected through a questionnaire and a comprehensive interview of teachers (N = 18), school principals (N = 15), and district office staff (N = 10).

The questionnaire and interview were constructed to gather information ranging from attitude toward tests and testing through sophistication in use of tests and measurement terms. Although the project reported here dealt most specifically with testing as a result of the state mandated California Assessment Program, important insights were gained regarding general testing practices.

Results and Discussion

Attitude

As the data were analyzed it became apparent that responses across groups were directly related to educational roles. That is, as individuals moved farther away from teaching students in the classroom, attitudes, expectations and needs regarding tests differed. It is these differences which appear to determine why tests and test results are used in certain ways.

Attitude toward tests and testing is one area where a difference across groups was noted. While all three groups thought that in the ideal situation tests could be somewhat to extremely useful, district staff and principals (100%) leaned more toward this viewpoint than did teachers (84%). Further, while district staff (100%) felt that

standardized tests should be administered once or twice a year, teachers (89%) and principals (87%) were slightly less agreeable. More indicative of attitude, however, are teachers' (11%) and principals' (13%) comments that standardized tests should never be administered. No district staff respondent voiced this opinion.

The above comments reflect teachers' views that tests such as the CAP do not yield direct instructional benefits for individual students in their classrooms. Across the three groups, there is a difference in perception regarding the use in which tests should be put. While 79% of the teachers interviewed felt that the primary purpose of testing should be for individual diagnosis, only 17% of the principals and 33% of district staff concurred. Thus, teachers are most concerned about the individual students in his/her classroom at his/her classroom period. District staff and principals are more global in their feelings toward tests, noting that district-wide assessment to measure group gains is a valuable function of standardized measurement.

On questions designed to determine test preference in different situations, respondents were asked to rank the suitability of standardized, criterion referenced, and teacher made assessment instrument. For testing on a school wide basis, teachers preferred their own teacher-made or criterion referenced tests to the standardized measures preferred by principals and district office staff. All three groups agreed that standardized tests were the best way to measure reading achievement on a state-wide basis. For testing a class of students, teachers ranked their own, teacher-made tests first, followed by

criterion-referenced measures. Standardized tests ranked last. Principals, however, ranked teacher made tests last in order of preference for testing on a class wide basis. It seems, therefore, that teachers place most faith in their own measures, closely followed by criterion-referenced assessments and place least faith in standardized testing instruments.

Perhaps the above attitudes have a basis in sophistication to testing techniques and terminology. On questions designed to provide insight as to knowledge of measurement terms and techniques, teachers appeared least knowledgeable. District staff were most knowledgeable. This again reflects the different educational responsibilities of the groups in our sample.

In short, teachers were generally less favorable toward tests and testing than were the other two groups. Perhaps this is because they are unclear as to the specific purpose for the CAP testing program, although this was true across groups. No one had ever seen a published statement of the CAP goals, nor could anyone state what the goals were. When subjects were shown the five CAP goals, and were asked to rank these in terms of how well they were being met, teachers were again substantially less favorable than were principals or district staff (Ruddell & Kinzer, 1981). A majority (56%) of teachers felt that the goal "To discover special strengths and weaknesses of California school children...to aid in improvement of school programs" was not being met, while a majority of principals (85%) and district office staff (56%) disagreed. This pattern held fairly constant through the five stated

goals of the CAP (Ruddell & Kinzer, 1981). A general dissatisfaction across groups, however, was seen in response to the goal "to provide the public with evidence of effectiveness of the educational system in teaching basic skills". Sixty-eight percent of the teachers, 54% of the principals, and 56% of the district staff personnel felt this goal was being met marginally or not at all.

Time Involvement

Ultimately it is the individual's perception of the use to which tests are put which determines the value which is placed upon them. Testing does take time--often time which would otherwise be used for instruction. Forty-four percent of the teachers stated that they needed approximately one hour to prepare for administration of the CAP, even though they found (84%) the test easy to administer when compared to other standardized measures. Fifty-three percent of the principals stated that they spent more than one hour to prepare for administration, while 50% of district staff respondents said that they spent more than two hours in preparation for administration of the CAP. This result again reflects the varying responsibilities of the three groups interviewed in our sample. District staff is responsible for coordinating distribution and administration throughout the district, principals for coordinating distribution and testing throughout their schools, while teachers use preparation time to pursue instructional manuals and to attend staff meetings before administering the tests in their classrooms.

Obviously, the role of the district staff which we interviewed was a major one with regard to testing, and to CAP in particular. Thus, although 90% of the teachers and principals said their staffs met to discuss CAP results, their role essentially ended following administration of the tests. District staff, however, is responsible for compiling the results, writing the report, forwarding this report to schools and school boards, and must make themselves generally available to discuss the findings. It is not surprising, therefore, that 60% of the district staff stated that CAP directly demanded four or more days of their time immediately after the test was administered, as well as taking additional time throughout the year. This is probably true of any district-wide, mandated testing program.

To repeat, testing is time-consuming. In order to justify this expenditure, educators must view the time as being well spent. Yet, when asked whether or not the CAP affects instruction in the classroom "A great deal, somewhat, marginally or not at all", the three groups were remarkably similar in their judgments. Thirty-nine percent of the teachers, 38% of the principals, and 50% of the district staff respondents felt that the CAP affects instruction marginally or not at all.

Usefulness and Impact

Teachers repeatedly commented that tests were not really disliked, but that there was simply too much testing done, which was felt to be

repetitive. The districts in our sample, in addition to the CAP, administered at least one other standardized test. The larger districts additionally used a district-developed measure. When coupled with everyday testing performed as part of any reading/language arts program, sundry teacher-made tests, and tests for minimum competency assessment, many of the respondents felt that there was too much time being spent testing when compared to instruction. Of published tests, the CTBS and SAT were the most commonly used in our sample.

The different responsibilities of the three groups we interviewed was also reflected in their perceptions regarding the impact of CAP. Through a study of the published CAP goals, we decided upon five areas where CAP might have the most impact: district curriculum planning, local school planning, teacher in-service need, classroom instruction, and budget decisions. Respondents were asked to rank these five areas in terms of CAP's impact on each. Impact upon district curriculum planning was ranked first by district staff and second by both teachers and principals. Planning at the local, school level was ranked first by principals, second by district staff, with teachers voicing no opinion. As a guide for in-service needs, CAP's impact was ranked third by teachers and district staff, fourth by principals. Impact upon the important area of classroom instruction was ranked fourth by teachers and district staff, third by principals. All groups agreed that CAP impacted on budget decisions.

The various roles in which the three groups interviewed find themselves in the educational setting marks a definite difference of

opinion in perceived usefulness of CAP and standardized testing generally. In answer to the question "How useful have the CAP results in reading been to you?" 34% of teachers found them somewhat to extremely useful, 28% had no opinion, and 39% found the results somewhat to completely useless. This finding varies dramatically from principals, none of whom reported that they felt the results useless, although 23% did find the results neither useful nor useless. Seventy-seven percent of the principals found the CAP results to be somewhat to extremely useful, while 80% of the district staff respondents found the results useful. The above findings are in sharp contrast to the 100% across groups, who felt that in an ideal situation tests could be somewhat to extremely useful. It is this perceived gap between the ideal and the actual situation which is of concern to us, especially given the amount of testing presently done, and the publicity which test results receive.

Recommendations and Conclusions

Teachers and principals felt strongly that individual students' scores were needed if maximum instructional benefit from CAP, and testing in general, is to be gained. A report of only class or grade-wide scores were thought inadequate by these respondents. An item analysis and a listing of students by item missed was strongly desired by teachers and principals. This concern for individual reporting of scores, however, was not shared by district staff, who are more involved with across-school comparisons.

Presently, reporting of individual scores is not possible, due to the matrix sampling procedure which is used by the CAP. This sampling procedure, it should be noted, was generally misunderstood by our respondents, and might be more clearly explained. Indeed, it seems that much of the concern we found regarding tests and CAP could be dispelled through a comprehensive rationale, and a clear statement of goals.

Across groups, there appeared to be little understanding of measurement terms, although this was more true of teachers than of either principals or district staff. All respondents preferred either percentile or grade equivalent scores for reporting of test results. The CAP, however, uses standardized and/or scaled scores to report results, and is moving away from the most-favored percentile. All but two of our respondents requested conversion tables so that percentiles could be readily accessed. Scaled scores were generally misunderstood. It is suggested that districts implement an in-service program on measurement terms and techniques. This would greatly help understanding and interpretation of test results.

All teachers and principals, and a majority of district staff called for CAP results much earlier in the year. Presently, CAP results from the previous year's (Spring testing) are disseminated in late Fall (usually November/December). Respondents felt this was too late to change instructional policy, and results in a possible two year time delay before instructional changes might be implemented. We felt that greatest instructional utility arises from the least time lag

between test administration and reporting of results. and suggest that CAP results be made available as soon as possible.

Finally, respondents felt that there was too much duplication of testing in schools. Principals, especially, felt that tests could be reviewed and a single, standardized test could be either chosen, constructed or modified in order to eliminate what they viewed as too much time spent on testing. Additionally, principals were concerned over the manner in which results were presented to the news media. They felt that results needed more careful explanation and structure in order to reduce misinterpretation. In conclusion, we note that the different educational roles and responsibilities of teachers, school administrators, and district office personnel influence their attitudes, and perceptions toward tests and test results. It further appears that each group can accept the other's needs and values providing testing goals are clearly explained and delineated. A needed next step is to interview legislators and parents in order to find how their perceptions, needs, and uses of test results match (or mismatch) the results presented in this report.

References

Ruddell, B., & Kinzer, C. Test preferences and competences of field educators, paper presented at the National Reading Conference, Dallas, 1981.

School District Testing

The Great Educational Earthquake of 1981: Reallocation of Resources in California High Schools to Meet Minimal Competency Requirements: Was the Reallocation Necessary?

Harry Singer

University of California, Riverside

Prepared while the author was supported, in part, by a California Policy Seminar grant, administered by the Institute of Governmental Studies (John Cummins, Director), University of California, Berkeley. The grant is to Harry Singer (University of California, Riverside), John McNeil and Merl Wittrock (University of California, Los Angeles), and Robert Ruddell (University of California, Berkeley) to study "Achievement Testing in Basic Skills." A final report on this grant is due June 30, 1982.

July 15, 1981

The Great Educational Earthquake of 1981: Reallocation of Resources in California High Schools to Meet Minimal Competency Requirements: Was the Reallocation Necessary?

Harry Singer

University of California, Riverside

The great educational earthquake of 1981 resulted in a rapid reallocation of resources at the high school level. This reallocation occurred to meet the educational requirements of some 100,000 students, about 30 percent of the tenth-grade class, who did not pass one or more proficiency tests in reading, writing, and arithmetic. These students then signed up in 1980-81 for one or more remedial courses in mathematics and English.[1]

Cost of the Reallocation of Resources

Simple arithmetic will show the magnitude of resources required to staff the courses which had to be offered in order to meet the Educational Code requirements (Sections 51215-17) that went into effect in 1977. The Code requires school districts to provide remedial courses for students who do not pass the district's proficiency tests. We know the 100,000 students took at least one remedial course. If we assume the 25 students are enrolled in a remedial class, then 4,000 remedial classes had to be offered to educate the 30 percent of the students who failed on one or more of the proficiency tests given in the tenth grade in 1979-80. If we further assume that five classes is a normal teaching load, then 800 full-time equivalent teachers taught these classes. If we also assume that these teachers each earn a salary of $15,000 per year, then the cost of the

instruction is 800 x $15,000, which equals $12,000,000. (That is almost as much money as California allocates from the general fund for all textbooks in grades K-4.) Because the State of California did not provide high schools with any additional funds for this instruction, school districts could only meet the new Educational Code requirement by taking instructional resources from the upper 70 percent of the students, those who had passed all the proficiency tests, and shifting them to the lower 30 percent of the students. English and mathematics teachers who had been teaching electives and advanced courses for college preparatory students found that they had to be reassigned to teaching proficiency classes. School districts also used funds to offer proficiency classes in summer school.

Trivialization of the Curriculum

Although no survey has yet been made on the effects that this great allocation of resources is having on faculty and students, the State Department of Education feared the potential impact would be "trivialization" of the high school curriculum.[2] We can define a trivial curriculum as one that does not add much to a student's general achievement. Another definition is that it is a curriculum without electives in fine arts or advanced courses in academic subjects. A trivial curriculum may adversely affect the school's achievement level in basic skills by not only eliminating the challenging advanced courses, but also by narrowing the curriculum.

To improve the average achievement of its students, a school ideally should increase the scores of all its students. However, students at the extremes of achievement affect the average level of achievement the most. Those at the lower end pull it down, those at the upper level lift it up.

Since students at the upper level tend to gain the most from instruction, school districts which want to improve results on such standardized tests as the California Assessment Program, the Comprehensive Test of Basic Skills, or the Iowa Tests of Educational Development should offer advanced courses to their students at the upper end of the achievement level to provide the necessary lift to the school's average level of achievement on standardized tests. But, if the school does not offer these students advanced courses, then it tends to preclude one way of improving the school's standing in achievement on standardized tests. In shifting resources from the upper 70 percent to the lower 30 percent, high schools are in effect robbing "Peter to pay Paul" and are thus operating on a myopic policy.

Impact of Remedial Courses

Nor are the schools necessarily benefiting the lower third by offering remedial courses. The institution of remedial courses in math and English reinstitutes a kind of tracking system in the high school. This system of grouping students for instruction was outlawed in Hobson vs. Hansen.[3] Ever since, high schools have allowed their students to select their own classes instead of having counselors program and assign them to classes based on their general mental ability or their general levels of achievement. Full tracking has not returned because counselors do not assign students to remedial courses, but students feel they must take these courses. This option takes them out of regular classes and puts them into special classes. These remedial classes run counter to the spirit of Public Law 94-142, which was formulated for special education, but its philosophy also applies

to low achieving students. PL 94-142 specifies that students in special education should be placed in the least restrictive educational environment. In other words, if students do not improve in basic skills and do not pass their proficiency tests by graduation time, they also suffer another stigma: they do not get a diploma at their high school commencement. Some school districts even bar them from participating in commencement exercises. The students are then not legally eligible for any position which requires a diploma. So, students feel compelled to sign up for remedial classes and suffer the stigma of being in classes segregated from the upper 70 percent of the high school students. Many of these segregated students are Black and Chicano. After all the efforts to integrate students, proficiency testing and remedial classes that accompany them are resegregating these students.

No evidence is available yet on the effects of remedial classes. That is, no knowledge has been provided that remedial classes are any more effective than regular English classes in improving students' reading and writing ability.

Instruction in Remedial Classes

We do know that the competency law did not provide any staff development funds for inservice education of teachers. Consequently, teachers did not go through any special training program to prepare for teaching the remedial courses. These teachers are probably unaware that students in these classes may fall into 10 diagnostic categories, with students in each category needing appropriate instruction. Teachers of the remedial classes may elect to offer a form of "individualized instruction" which is the poorest

kind of instruction: no instruction. This type of instruction consists of dittoed assignments, like workbook exercises, which give students written directions and then a set of problems to do. However, the teacher offers the students little or no instruction. Indeed, with 25 students in various assignments to be done during a 50-minute period, the teacher who is administering the "individualized instruction" only has two minutes per student, which is hardly enough time for adequate instruction for each student. Essentially, this instructional program is misnamed. It is not individualized instruction; instead, it is "bootstrap instruction" in which students must lift themselves up by their own bootstraps. That is, they must learn on their own from the dittoed assignments. They must read to find out what to do, try to do it, and if they make an error, diagnose the error themselves and correct it.

In contrast, it is possible to train teachers who could offer a more appropriate type of course. These teachers could not only diagnose and group students for needed instruction, but use a variety of input instruction. For example, in one type of input instruction, the teacher would start a lesson by explaining and demonstrating how to do a new exercise or a problem, then call upon students to see if they understand the new instruction, and finally, when the teacher is satisfied students understand, only then assigning practice problems in the text, workbook, or dittoed exercise sheets! They could also use teaching strategies for all the students in the class, including a directed reading activity,[4] reading guides,[5] and reading acquisition in content materials.[6] These strategies also enable teachers to instruct all students in regular content area classes who are beyond the elementary stage of reading acquisition.[7,8]

Through such instruction, teachers may be able to help all students improve in their basic skills enough to pass the school's proficiency tests. Moreover, these classes would not stigmatize and segregate students, nor would they soften the curriculum. Whether this approach results not only in improvement on proficiency tests but also on achievement in basic skills, we know that these classes will not require reallocation of resources, but only training of teachers in strategies to handle a wide range of individual differences in achievement.

Normal Range of Achievement

The normal range of individual student differences at the tenth-grade level consists at one extreme of students whose reading ability is fifth-grade level in general reading ability, and at the other extreme students whose general reading ability is comparable to the average level of reading ability of a second-year college student.[9] This range of ability occurs when all students are achieving up to the limits of their general mental ability.

Only recently have high school teachers had this full range of reading abilities in their classrooms. In the 1950's only 55 percent of the students in the ninth grade remained through high school. In the 1970's this figure had increased to 85 percent.[10] Consequently, high schools during the first part of the century were highly selective institutions. Many students in the early 1900's thought their education was complete at the end of the sixth or eighth grade, In fact, the junior high school was created to serve as a transitional unit that would encourage students to go on to high school, particularly students with high achievement. Those

with low achievement continued to drop out and enter the job market. Gradually, the holding power of the high school climbed and had a dramatic spurt in the 1960's and 1970's. But high school teachers were unprepared to teach students with such a wide range of abilities. In self-defense, they tended to voice the opinion that elementary and junior high schools should have brought all these students up to grade level before they reached high school.[11] But the concept of grade level represents an average; by definition, 50 percent of the students will be below average and 50 percent will be above. What the high school teachers meant is that only the upper 50 percent of the junior high school graduates should go on to high school.

The proficiency test is in effect a movement in this direction. Perhaps this movement was not intended by those who favored the establishment of the minimal competency law, such as legislators Hart and McAllister, and minority group leaders. Their intention may have been to reallocate school resources for low achievers and hold schools accountable for instructing them, not just promoting them regardless of level of achievement as schools had been doing under the social promotion policy in effect since the 1950's.[12] Others may simply want the high school diploma to certify that recipients have a minimal competency in basic skills.

Legal Requirements for Proficiency Tests

However, the law does not require schools to use a proficiency test as a screening device for high school students. Indeed, California's minimal competency law, in contrast with New York's, allows each school district to select or construct its own competency test and determine its

own score for passing the test. Under the law, California's State Board of Education may only furnish guidelines and examples for proficiency tests. In fact, the law (AB 65) expressly forbids the State Department of Education from contructing and administering a proficiency test. Therefore, instead of the State Department of Education and ultimately legislators having the responsibility for the proficiency test and its consequences, the legislature passed the responsibility on to the local school districts. In contrast, the New York State legislature authorized one state proficiency test and the same passing score for all students; hence the New York legislators take responsibility for the validity of the test, establishing the passing score for the proficiency test, and for assuring the public that all students in New York State who have been awarded a high school diploma have attained or excelled the same minimal level of competency. In contrast, a wide variety of tests were selected or constructed in California. Some 22 percent of the school districts in California selected a proficiency test, such as the Educational Testing Service's or McGraw-Hill Publishing Company's SHARP Test (constructed and developed by the Los Angeles Unified School District) or else used a commercial pool of items to construct a test. The commercial tests have about an eighth- or ninth-grade level of difficulty.[13] Their cutting score is set at a level where only 70 percent of the ninth graders pass the test. School districts which use these or comparable tests must then provide remedial courses for the students who fail one or more sections of the test. To do so, high school resources must be reallocated.

Although our society has not had a consensus on a standard for literacy, we have had some rules-of-thumb. The U.S. Bureau of Census uses as its

criterion of literacy completion of the sixth grade or only a positive answer to the question of whether a person has the ability to read and write a simple message in any language.[14] But school boards which have adopted an eighth- or ninth-grade level of difficulty for a test and a passing score of the 10th percentile or higher have raised, perhaps inadvertently, the minimal standard for literacy to a much higher level, higher than it has been in any society at any time! When viewed from an historical perspective, the change is dramatic. In England and France over 200 years ago, a person only had to read and write his or her own name to be considered literate.[15] In the United States 20 percent of the citizens could not reach this criterion in 1870. By 1978, less than one percent of citizens 14 years or older were illiterate according to the same criterion.[16] But with the advent of the new standard for minimal competency in basic skills, those students who do not pass the proficiency test are now deemed to be illiterate until they do pass it.

The new standard for literacy involves for the first time (1) in most states, a formal test of literacy, and (2) for the general public, a change in the criterion of literacy from only the ability to read familiar material aloud to an examiner's satisfaction, or attendance at school for six years or more, or an affirmative response to an inquiry on whether a person knows how to read and write a simple message to a criterion of <u>demonstrated</u> ability to <u>understand</u> <u>an</u> <u>unfamiliar</u> <u>text</u> and use that information in new situations.[17] In California, the number of tenth graders who are illiterate according to the new criterion is 30 percent. We do not yet know how many of them will pass by the twelfth grade, but we do know that considerable resources are being reallocated and marshalled

during the academic year and particularly in summer session classes to help them reach the new level of literacy. Only now are we beginning to realize the consequences and costs of adopting a minimum competency test. Thus, without conscious intent, our society has raised the minimal standard for literacy by two or three grade levels.

Of course, not all school districts set the minimal competency at such a high level. One school district established minimal competency at the sixth-grade level. At this level, the students have learned to read the words they hear in everyday life, write sentences for language use in ordinary conversation, and do all basic arithmetic computations. Those students who do not pass this level of proficiency can take additional courses at the junior high school and can probably pass the test before they reach high school. Consequently, this school district's high school will not have to shift instructional resources to the lower 30 percent by setting up remedial courses; it will not have to develop a quasi-tracking program, resegregate students, stigmatize them, and deny anyone who completes high school from receiving a diploma. Nor will its students want to drop out of school rather than face embarrassment at graduation.

Effects of High Minimal Competency Standards

Why did other school districts adopt a higher level of achievement? Probably the reasons are multiple. Some districts may have opted for reallocation of resources as a way of improving the minimal competency of their students in the interest of social justice and equality of opportunity. Other school districts may have confused minimal competency with minimal requirements for high school graduation and in so doing, established proficiency tests and standards for passing them at a relatively high level. One superintendent reported his Lions Club found his school

district's proficiency test had some embarrassingly difficult questions when he asked them to examine the test during a luncheon talk.

No school district has made any study to determine what adults have to know and be able to do in reading, writing, and arithmetic to be minimally competent! However, other agencies have produced some information on literacy; for example, we have some knowledge of the literacy requirements for Army jobs and for a wide variety of other occupations.[18,19] But we recognize that determination of literacy or minimal competency in reading is a complex issue. It involves not only the reading ability required for a particular job but also characteristics within the reader and the reading task, such as the aspirations of the reader to read certain types of material (e.g., high school or college level, literary and technical) and variations in the readability of the material.[20] Indeed, readability involves an interaction between the resources within the reader (prior knowledge, reasoning capacities, vocabulary ability, desire to know and understand),[21] and properties of the text.[22]

Despite the complexities involved in determining minimal competency and lack of evidence on it, school districts have nevertheless adopted or selected proficiency tests that are irreversibly affecting the certification of an estimated seven percent of California high school seniors. These seniors, some 21,000 students, were not expected to receive a diploma in 1981.[23] Yet the Harris survey of literacy indicated only one percent of the 16-year-old students across the nation were illiterate; correctly answer questions of Harris's life-skill type of reading materials.[24] This finding suggests that high school proficiency tests are more difficult than the Harris materials.

School Board Concerns

A study of minimal adult competency is necessary before schools can justify the validity of their proficiency tests. When certification is the issue, then courts will eventually decide whether the legal rights of students denied a high school diploma have been violated.[25] In anticipation of this consequence, school boards would be wise to ask their directors of research and evaluation to determine the reliability and validity of their proficiency tests. They should specifically determine whether the tests have curricular and instructional validity; that is, whether the test's objectives are part of the curriculum and whether the contents and processes assessed on the test are actually taught in the classroom. Otherwise, school boards may find themselves in court defending themselves from suits brought by students denied a high school diploma.

School boards also have to be concerned lest students and parents in the upper 70 percent react negatively to the reallocation of resources. The Educational Code requires high schools to prepare students for entry to colleges and universities. We already know that California students take fewer academic courses, for less time, get higher grades, and expect more advanced placement at the colleges and universities they plan to enter than do students throughout the nation who take the College Board Examination.[26] When California students enter colleges and universities, they are not only likely to find they do not get advanced placement, but also do not earn as high grades. This shock will occur particularly for those students who enter colleges and universities outside of California. Colleges and universities in California have already expressed concern about the drop in high school achievement.[27] Indeed, the legislature should assess the adequacy of the high school's college and university preparatory

program and monitor whether achievement in it is dropping, particularly in the subjects required for admission to the University of California. If so, then it will have to eventually deal with the issue that schools are not preparing their academically talented youth adequately for society's future needs.

What should board members do? The most obvious solution is to reconsider their district's proficiency test. They can also ask the legislature to fund a study to determine what minimal competency is in the adult community. Until the legislature acts, school boards can establish minimal competency at a level where 70 percent of the sixth-grade class can pass the proficiency test. School boards should also make sure that the content and processes assessed on the test represent samples of the curriculum that are actually taught in the classroom. If school boards adopt this strategy, then we can expect quick recovery from the great educational earthquake of 1981. If they do not act, then the aftershocks are likely to continue for some time.

References

[1]Proficiency Assessment in California: A Status Report. Sacramento, Calif.: State Board of Education, 1980.

[2]Ibid.

[3]Hobson V. Hansen, 269 F. Supp. 401 (DDC 1967).

[4]Stauffer, Russell G. Directing Reading Maturity as a Cognitive Process. New York: Harper and Row, 1969.

[5]Herber, Harold. Teaching Reading in the Content Areas, Second Edition. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

[6]Singer, Harry and Donlan, Dan. Reading and Learning from Text. Boston: Little, Brown, 1980.

[7]Ibid.

[8]Singer, Harry. IQ is and is not Related to Reading. In S. Wanat (Ed.), Issues in Evaluating Reading. Arlington, Va.: Center for Applied Linguistics, 1977.

[9]Goodlad, John I. School, Curriculum, and the Individual. Waltham, Mass.: Blaisdell (Ginn), 1966.

[10]Jencks, Christopher S. Effects of high schools on their students. Harvard Educational Review, 1975, 45, 273-324.

[11]Singer, Harry. Attitudes towards reading and learning from text. In M. L. Kamil and A. J. Moe (Eds.), Reading Research: Studies and Applications. 28th Yearbook of the National Reading Conference. West Lafayette, Indiana: The National Reading Conference, 1979.

[12]Caswell, Hollis S. and Foshay, A. Wellesley. Education in the Elementary School. New York: American Book, 1950.

[13]Singer, Harry; McNeil, John; Ruddell, Robert B. and Wittrock, Merl C. Testing Achievement in Basic Skills (Final Report). Berkeley, Calif.: California Policy Seminar, Institute of Governmental Relations. (in preparation)

[14]Fisher, Donald L. Functional Literacy and the Schools. Washington, D.C.: The National Institute of Education, 1978.

[15]Resnick, Daniel P. and Resnick, Lauren B. The nature of literacy: An historical exploration. Harvard Educational Review, 1977, 47, No. 3, 370-385.

[16]Fisher, op. cit.

[17]Resnick and Resnick point out that in France's elite technical schools in the 19th and 20th centuries, sons of the aristocracy, bourgeoisie, and some able sons of the poor learned mathematics as the key to effective reasoning and acquired technical literacy. But in the United States, it was not until the 1920's with the advent of emphasis on silent reading and the invention of reading tests did the "ability to understand an unfamiliar text rather than simply declaiming a familiar one [became] the accepted goal of instruction" (p. 382). Furthermore, this "newer standard, previously applied only to elite institutions, required the ability to gain information from reading and use that information in new contexts" (p. 382)

[18]Sticht, T. et al. Project REALISTIC: Evaluation and Modification of REAding, LIStening, and ArithmeTIC Needs in Military Jobs Having Civilian Counterparts. HumRRO Professional 19-71. Human Resources Research Organization, Alexandria, Virginia, Spetember 1971.

[19]Sticht, T. and McFann, H. H. Reading requirements for career entry. In D. M. Nielsen and H. H. Hjelm (Eds.), Reading and Career Education. Newark, Delaware: International Reading Association, 1975.

[20]Bormuth, J. Reading literacy: Its definition and assessment. In J. B. Carroll and J. S. Chall (Eds.), Toward a Literate Society. New York: McGraw Hill, 1975.

[21]Singer, H. The substrata-factor theory of reading: Its history and conceptual relationship to interaction theory. In L. Gentile and M. Kamil (Eds.), Reading Research Revisited. Columbus, Ohio: Merrill. (in press)

[22]Kintsch, W. and Van Dijk, Teun A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

[23]Op. cit.: Proficiency Assessment in California: A Status Report.

[24]Harris, Louis, and Associates. Survival literacy study. Congressional Record, 1970, November 18, E 9719-9723.

[25]McClung, M. S. Legal implications of intelligence and competency testing, Center of Law and Education, Harvard University (Summary). Complete article is "Testing for Intelligence and Competence: Legal Implications" and is available from Gutman Library, 6 Appian Way, Cambridge, Mass., 02138, 1979.

[26]McCandless, Sam. The academic experience of California's college bound high school seniors as reported by the students themselves in the 1978-79 Admission Testing Program, summarized in the California State Board of Education: Summary of Initial Discussions on Secondary Curriculum, January 10, 1980 (Revised August 13, 1980). In: Report to the Policy Committee on the University of California's Activities to Assist Underprepared Students, Part 1. Berkeley, Calif.: University of California, March 1981, pp. 70-71.

Preference for Testing and Reporting of Test Results in Basic Skills: Acquisition vs. Ability to Gain Information from Text, Content-Bound and Content-Free

Harry Singer
University of California, Riverside

and

Mariam Jean Dreher
University of Maryland,
College Park

January 1982

Preference for Testing and Reporting of Test Results in Basic Skills: Acquisition vs. Ability to Gain Information from Text, Content-Bound and Content-Free

Abstract

Four ways of testing and reporting test results were ranked ordered from most to least preferred by a group of educators enrolled in an administration course at the University of California, Riverside. They also completed Likert-type Reaction Scales for each of their accepted and repeated choices. The four choices were based on two tests: (1) standardized reading test administered in grade 1.8 and readministered in grade 2.8; both test results were interpreted according to grade 1.8 norms; the graphed result was called self-comparison. (2) A second standardized test, normed referenced for and administered in grade 2.8; a graph of the 1.8 and 2.8; a graph of the 1.8 and 2.8 results for the two different tests was labeled peer-comparison. Another factor was content-free vs. content-bound reporting of test results. The content-free graph of the results depicted only numerical information; the content-bound graph contained numerical information plus the most difficult paragraph and questions a student could successfully read on each test. The most preferred graph of test results contained self- and peer comparisons with content-bound. The least preferred graph consisted of peer-comparison only, content-free. Although responses to the 32 Likert Scale items revealed statistically significant differences between the two

choices, the educators expressed agreement with only 9 of these items. These items indicated they agreed that the combination of self and peer comparison testing and test results, content-bound is helpful to parents and teachers, informs them on student improvement, indicates how a student compares with others, reveals the student could read more difficult materials after another year of instruction, informs the teacher on student progress, provides teachers with information they did not already know, and thus the tests and test results provide them with useful information. The study was completely replicated on a similar group of educators at the University of Maryland. Hence, the ambivalent and somewhat negative attitudes of educators towards tests and test results can be reliably shifted in the direction of more favorable responses by tests and test results that are more diagnostic and content-bound. Consequently, the results call for a change in current practices in testing and reporting of test results in reading, and suggest the need for differentiation in the assessment component of a theory of instruction in basic skills.

Dr. Harry Singer
School of Education
University of California
Riverside, CA 92521
714-787-5228 (office)
714-686-9195 (home

Dr. Mariam Jean Dreher
Reading Center
University of Maryland
College Park, MD 20742

Preference for Testing and Reporting of Test Results in Basic Skills: Acquisition vs. Ability to Gain Information from Text, Content-Bound and Content-Free

Harry Singer[1]
University of California,
Riverside

Mariam Jean Dreher
University of Maryland
College Park

Standardized tests have enjoyed steadily increasing sales in the United States, from $7,000,000 in 1948 to $42,000,000 in 1976 (only $24,000,000 when corrected for inflation since 1967) (Resnick, 1981). The receptive atmosphere for the tests stems historically from American ideals for objective selection of talented students regardless of social origins, desire for efficient use of resources and rational placement and grouping of students so as to reduce the failure rate, and national standards for evaluating school achievement and thereby limiting autonomy of local school districts (Resnick, 1981).

Although parents, the press, and the public want more testing, school personnel are ambivalent about standardized testing. Central office administrators in 18 school systems (public, private, and parochial) in Western Pennsylvania were surveyed. The public school administrators think that principals and teachers benefit the most from the tests. Administrators in private schools believe that parents, students, and communities have the greater need for test results than do school personnel (Sproul & Zukow, 1981). When interviewed, 68 teachers in a suburban school district near Pittsburgh and a large urban school district reported that standardized tests, such as the Metropolitan, Stanford, and California Test of Basic Skills were useful in supplementing their information on students, but placed more credence in their own observation of children's performance in the classroom than on standardized test scores, particularly when test scores were lower than expected. When they were higher, teachers reexamined their assumptions about students. The teachers stated a need for tests that are diagnostic, closely match the curriculum and instruction, and yield the information quickly (Salmon-Cox, 1981). Administrators and teachers in Ireland where testing was introduced for experimental purposes arrived at similar opinions after four years of experience with standardized tests (Madaus, 1981).

Despite the criticism of tests from professionals, civil rights workers, and consumer advocates, parents, business people, and legislators want more testing (Madaus, 1981). In a national Gallup poll conducted in 1979 parents overwhelmingly responded that standardized test scores were "useful" or "somewhat useful" (Lerner,

1979). Although some organizations, such as the Council for Exceptional Children and the National Education Association have called for a moratorium on testing, a 1978 survey found that only 16 percent of 3500 teachers in a national sample agreed with this position (Stetz & Beck, 1980). However, California teachers have an ambivalent attitude towards standardized tests. Ninety-two percent did not favor their use, but 84 percent reported they would be unable to teach effectively without them (Takeuchi, 1977). Another national poll, conducted by the National Education Association, found that 84 percent of the teachers in the survey felt that standardized test scores could be misleading and biased but 90 percent nevertheless think that standardized tests are appropriate when not used in isolation (NEA, 1978).

All of the studies cited have been based on current test practices which usually involve the reporting of only numerical test data such as percentiles and stanines, particularly when test results appear in the newspapers (Green, 1975). Although some experimentation in testing has been conducted since Goslin's (1967) well-known study, such as Madaus's (1981) and those reported in a recent conference on testing (Tyler & White, 1979), a search of the current research literature failed to disclose any experimentation on ways of testing and reporting test results. Perhaps the controversy between professionals and the public can at least be reduced and perhaps the ambivalence of educators towards tests can be more or less resolved by testing and reporting of test results that more closely satisfy their request for tests that are more diagnostic and sensitive to their

curriculum and instruction (Singer, 1977a,b; Singer, McNeil, Ruddell, and Wittrock, 1982). Certainly some form of testing is not only a necessary part of practice, but also an essential component in a theory of instruction.

A Theory of Instruction

A theory of instruction includes instructional input, text or the means to a goal, and the goal itself. When a text is the means to the goal, then the goal is usually comprehension test. Learning is more likely to occur when there is coherence among these components, that is, when the instructional objectives, such as answers to teacher-based questions, can be obtained through the text and the comprehension test assesses students on the questions posed (Rothkopf, 1982). However, this sequence should also include feedback to the teacher and the student on degree of attainment of the goal so that the teacher can then determine whether the instructional objective has been attained. If the student has only partially attained the goal, then further instruction and practice may be necessary. The basic assumption underlying this model is that the student has the ability to attain the goal.

Students throughout the normal range of mental ability can attain the goals of all the basic skills, such as learning to read, compute, and write (Bloom, 1971), provided that they have instruction that is properly sequenced and adequate time for learning the tasks (Carroll, 1963). Under these conditions, IQ has a decreasing correlation with acquisition of basic skills as a group of students progresses through school (Singer, 1977).

Initially acquisition of basic skills is a goal of instruction, but as students acquire these skills, each of the basic skills becomes a means to the attainment of further knowledge. For example, reading is used to gain information from text, computations for solving problems in mathematics, and writing to express ideas and create all kinds of prose. Unlike the first phase of reading instruction, which consists of the acquisition of a set of basic skills that are finite, the second phase of reading instruction, the use of these skills for learning from text, is infinite. As a group of students progresses through school, individual differences in knowledge gained from text increases because the knowledge to be gained is infinite and individuals differ in rate of learning this knowledge.

Performance in use of basic skills or ability to gain knowledge from text is assessed by criteria that increasingly resemble functions assessed on intelligence tests. For example, reading achievement tests at successive grade levels have more difficult vocabulary, more complex syntax, more abstract concepts, require more complex inferences, and assume more prior knowledge on the part of the reader. Intelligence tests also increase in difficulty on similar variables. Under these conditions, the correlation between achievement and IQ increases as a group of students progress through school (Singer, 1977).

Acquisition and use of basic skills can be confounded when standardized tests, norm-referenced for each grade level, are administered to a group of students as they progress through school. For example, improvement in reading acquisition and use of reading for

gaining information from text are both reflected in test scores at successive grade levels. These scores are frequently interpreted as percentile rankings. This type of interpretation of test scores only informs consumers on the relative standing of students, not on progress made since administration of the previous test. Although grade equivalent scores would indicate that progress has been made, these scores still do not differentiate acquisition from application of basic skills.

To assess any kind of acquisition or learning of a skill, it is necessary to keep constant the task or the goal to be learned. Then, for learning to occur, there must be active reaction of the learner to relevant instructional situations with reinforcement or knowledge of results (Hilgard and Bower, 1966). An assessment task which is appropriate provides knowledge of the results. Based on feedback from the assessment, further instruction with some modification may ensue. Then the same assessment task is readministered and progress noted. In short to assess learning, the paradigm consists of teaching, testing, feeding back and acting on test results, reteaching, and then testing again on the same task. Through such a procedure we can differentiate acquisition of a basic skill from its application.

Differentiating the Basic Skill of Reading

Although learning to read cannot be completely separated from use of reading to gain information in the instructional paradigm, we can approximate assessment of learning to read by repeating the same reading comprehension test after each period of instruction until students attain, say, a criterion of 85 per cent correct comprehension

responses or at least 85th percentile ranking on the norms for the readministered test. When they have attained this criterion we can infer that they have made progress in learning (a) correct responses to printed words, (b) use of their language background, knowledge, and reasoning ability in word identification and in comprehension and (c) integration of (a) and (b). We call this type of assessment repeated or self-comparison testing. We can continue this type of testing until students are able to read and comprehend materials that represent a consensual criterion that the students had learned to read. Emphasis in instruction and assessment then can shift to ability to gain information from text.

When students improve in use of reading to gain information as reflected in reading achievement on tests standardized and norm-referenced for each grade, we can infer that they are not only gaining in reading acquisition but also in their own resources for interacting with text. These resources include improvement in language ability, background knowledge, and reasoning processes. We call this type of assessment peer-comparison testing.

We can contrast self-comparison testing with typical peer-comparison testing. Elsewhere we have used a hypothetical distribution of results to depict what we might expect to obtain by testing a representative class of students on both types of testing (Singer, 1977a, b). The distributions for repeated testing showed decreasing variability while peer-comparison testing revealed increasing variability as a group of students progressed through the grades. We inferred from the repeated testing results that the class was

developing rapidly in acquisition of reading ability, and becoming more alike in this phase of reading. Also, we could infer that the group was becoming more heterogeneous on the peer-comparison measures of reading attainment. In general, the variability for such peer comparisons is equal to two-thirds of the group's median age. (For an actual distribution of reading achievement for grades 2-8, see Cook, 1951.) Thus, we can differentiate reading acquisition from ability to use reading in interaction with text to attain comprehension for increasingly more difficult tasks.

Content Free vs Content-Bound Test Results

Test results, if reported in the usual way, by reporting only percentile ranking, do not inform us on what the students know and can do at any point in the testing program. When releasing test results to the newspapers, when sending test results home, and frequently when reporting test results to pupils school personnel, and board members, schools typically report only numerical scores, such as percentiles or stanines (Green, 1975). We call such reporting of test results as content-free. To provide content-bound information, schools would also have to supply information on the content from which the results had been obtained. For example, the schools might have reported to the newspaper the most difficult paragraph that students had to read successfully to attain at least a 50th percentile ranking on each of the peer comparison tests, or they could have provided parents with the most difficult paragraph at each grade level that their children had read correctly. If school staff had done so, then the students, their parents, and other school personnel would know not only the

students' scores but also from what content the students had successfully read and gained information.

Up to this point, we have explained the rationale and the procedure for differentiating acquisition of reading-ability on a fixed measure from ability to interact with text on increasingly more difficult measures. We also contrasted content-free with content-bound reports. Thus, we have discussed two types of testing (self-comparison and peer-comparison) and two ways of reporting test results (content-bound and content-free). We then used these four ways of testing and reporting test results for depicting a second grade student's test results. This student had taken a standardized reading achievement test, norm referenced for grade 1.8 at the end of grade 1 and took the same test a year later at the end of grade 2.8, and at the same time also took a standardized test norm-referenced for grade 2.8. We graphed these results in two ways. The first way of testing and reporting, peer comparison, Figure 1, shows only the results for the standardized, norm referenced tests administered at the ends of grade 1.8 and 2.8. This type of testing and reporting is the usual way schools test and report test results. The second way, self-comparison testing, as shown by the solid line in Figure 2, indicates the results of the repeated testing using the same test and the same norms for the grade in which the test was originally administered.

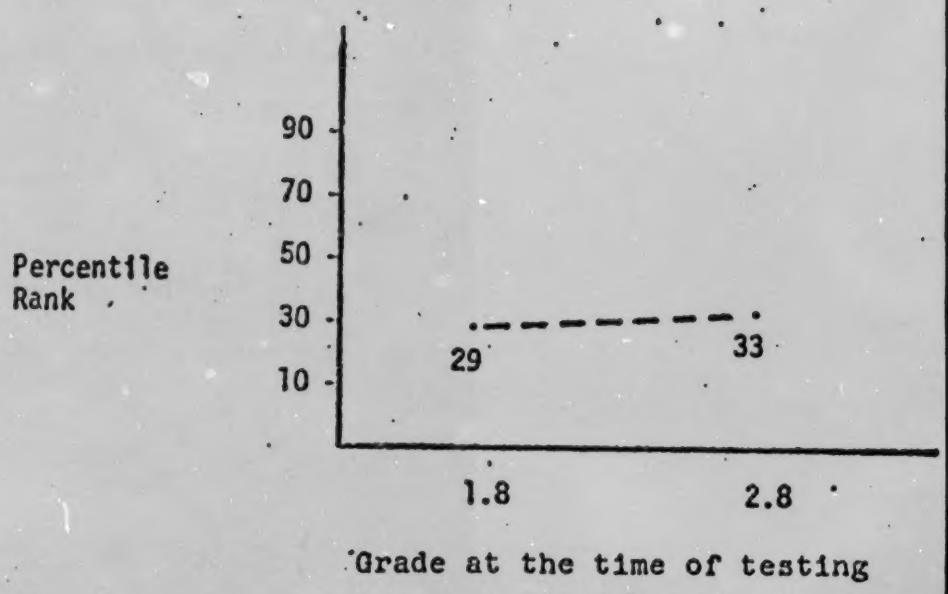


Fig. 1. This graph shows the student's percentile rank on reading comprehension on the Comprehensive Test of Basic Skills. At the end of first grade (1.8) this student was tested using the first-grade test. He scored at the 29th percentile, getting 12 out of 24 items correct. Then at the end of second grade (2.8) the student was tested with the second-grade test. He scored at the 33rd percentile, getting 11 out of 18 items correct.

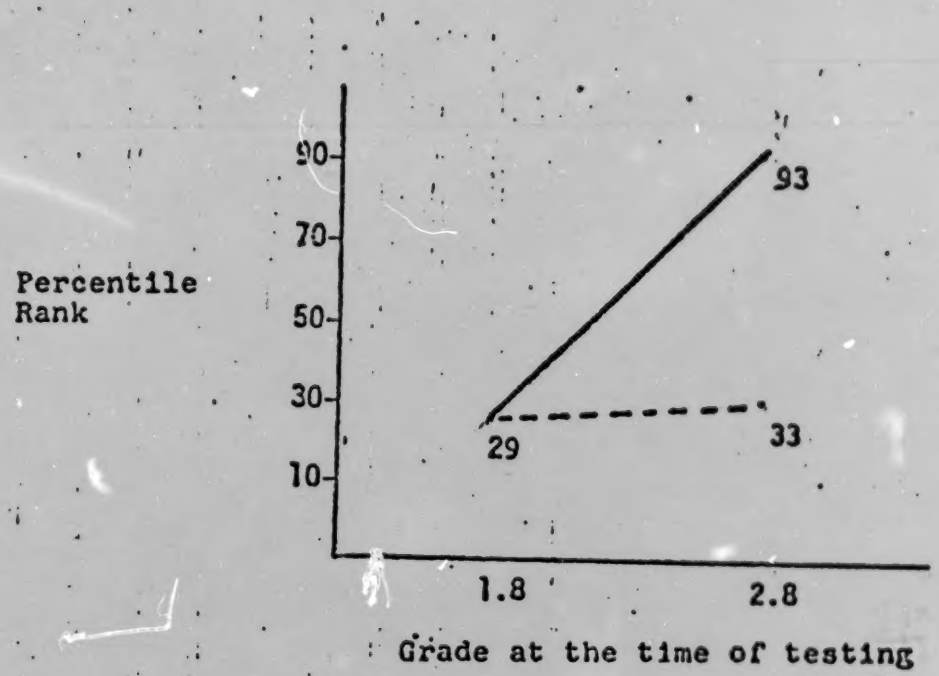


Fig. 2. This graph shows the student's percentile rank on reading comprehension on the Comprehensive Test of the Basic Skills.

At the end of first grade (1.8) this student was tested using the first-grade test. He scored at the 29th percentile, getting 12 out of 24 items correct.

At the end of second grade (2.8) the student was given two tests. First, he was retested with the same test he had taken in first grade. This retest shows that when compared to students in first grade, he now scores at the 93rd percentile, getting 24 out of 24 items correct. Second, he was also tested with the second-grade test. He scored at the 33rd percentile, getting 11 out of 18 items correct.

The dashed line in Figure 2 is the same information reported in Figure 1: the percentile rankings for standardized, norm referenced tests administered at the ends of grades 1.8 and 2.8. We repeated these two ways of reporting test information with the addition of content information in Figures 3 and 4, respectively. The content information indicates the most difficult paragraph successfully read in each test.

Purpose

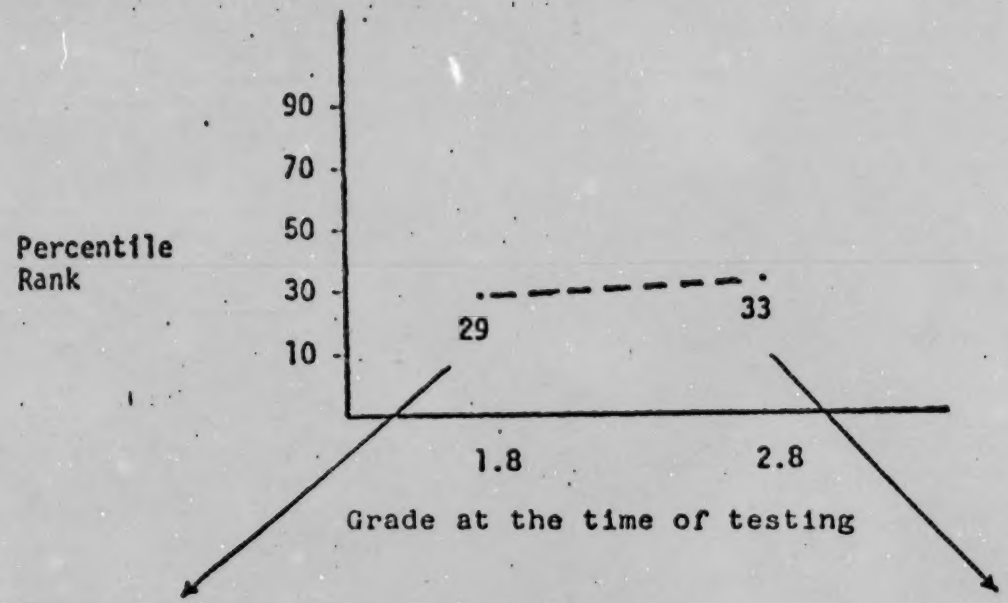
We then wanted to know whether experienced school personnel would prefer one type of testing and reporting of test results over another. If they did, we wanted to know why. We constructed an experimental design to answer these questions.

Experimental Design

The experiment was conducted in two parts. In the first part, experienced school personnel ranked the four categories of test results, Figures 1 to 4, from most to least preferred.[2] Although order of presentation of the categories probably would not have influenced the rankings, they were nevertheless presented in a random order to each person. Each person then filled out a Likert-type reaction scale consisting of 32 items for their most preferred and for their least preferred category of test result. This part of the study was done in counter-balanced order. Each person indicated on each reaction scale his or her most and least preferred type of test result.

Subjects

A total of 26 subjects participated in the University of California study. They were students enrolled in an administration credential course who had an average of 5.1 years teaching experience;



is is the hardest test passage the student
s able to read correctly on the first-grade
st at the end of first grade.

This is one book. It is in a box.

This is the hardest test passage the student was able to read correctly on the second-grade test at the end of second grade.

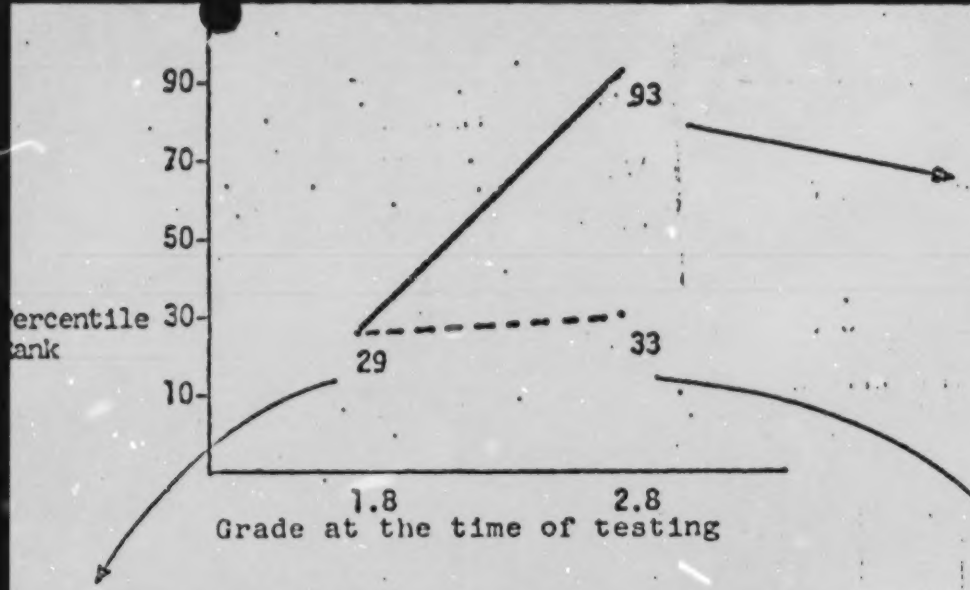
It was Ann's birthday. Mr. Parks had a surprise for Ann. "Here is the box of candy," he said. "Thank you, Father," said Ann, "I will put it on my toy box. I must go to school now, Betty Smith is waiting for me."

Frisky barked when he saw Ann go away. Then he ran to the toy box and got the candy. Frisky had a good time.

What did Frisky do?
- O ran away
- O ate the candy
- O went to school
- O brought Ann a gift

What was Father's last name?
- O Brown
- O Frisky
- O Smith
- O Parks

Figure 3. This graph shows the student's percentile rank on reading comprehension on the Comprehensive Test of Basic Skills. At the end of first grade (1.8) this student was tested using the first-grade test. He scored at the 29th percentile, getting 12 out of 24 items correct. Then at the end of second grade (2.8) the student was tested with the second-grade test. He scored at the 33rd percentile, getting 11 out of 18 items correct. Example test passages are shown under the graph.



This is the hardest test passage the student was able to read correctly on the first-grade test at the end of first grade.

This is one book. It is in a box.

This is the hardest test passage the student read correctly on the first-grade test at the end of second grade. (Note: He got all the passages correct.)

Mrs. Brown is in the city with a balloon.

This is the hardest test passage the student could read correctly on the second-grade test at the end of second grade.

It was Ann's birthday. Mr. Parks had a surprise for Ann. "Here is the box of candy," he said. "Thank you, Father," said Ann, "I will put it on my toy box. I must go to school now, Betty Smith is waiting for me."

Frisky barked when he saw Ann go away. Then he ran to the toy box and got the candy. Frisky had a good time.

What did Frisky do?
- O ran away
- ✱ ate the candy
- O went to school

What was Father's last name?
- O Brown
- O Frisky
- O Smith

Figure 4. This graph shows the student's percentile rank on reading comprehension on the Comprehensive Test of the Basic Skills.

At the end of first grade (1.8) this student was tested using the first-grade test. He scored at the 29th percentile, getting 12 out of 24 items correct.

At the end of second grade (2.8) the student was given two tests. First, he was retested with the same test he had taken in first grade. This retest shows that when compared to students in first grade, he now scores at the 93rd percentile, getting 24 out of 24 items correct. Second, he was also tested with the second-grade test. He scored at the 33rd percentile, getting 11 out of 18 items correct.

Example test passages are shown next to and underneath the graph.

eight taught at the elementary, four at the junior high, and five at the high school level. Five were administrators at the elementary or high school levels and two had some administrative position. The modal number of courses in tests and measurements taken by the group was one. Two subjects did not complete the information inventory.

The study was repeated at the University of Maryland on 18 students enrolled in a graduate course in research in reading. Thirteen of the students had seven or more years of teaching experience; the others all had from four to six years of teaching experience. Nine were currently teaching at the preschool through elementary level, two at the junior high, and one at the high school level. Two were principals of elementary schools and one had another administrative position. Three subjects neglected to indicate their current positions but were known to be experienced teachers. The modal number of courses in tests and measurements taken by the group was two.

Directions

1. Although some school personnel do not want to administer any standardized test, nevertheless, standardized tests are often required. However, their results can be reported in ways that will make them more useful and appropriate to teachers, administrators, and parents. Attached are four types of reports on the same student at the end of grade 2. After you rad the reports, rank them in order from 1 to , with 1 being your most preferred and 4 your least preferred. Then record the symbol at the top of each report next to its rank number on the Rank Order Scale. This Scale is on the next page.

2. Keeping your most preferred report in mind, read and record your reaction to each statement on the first Test Result Scale. Then, record your reaction for your least preferred report on the second Test Result Scale.

Analysis of Results

Rank Order

The results of the rank ordering indicated that fifteen subjects preferred repeated testing, content-bound. Their least preferred choice was peer group comparisons only, content-free. The other subjects chose a variety of combinations. Three of them most preferred repeated testing, content-free, and least preferred peer comparison, content-free. Two preferred repeated testing, content-free, and least preferred repeated testing, content-bound. The remaining five each had a different combination of choices.

The significance of the choices was evaluated by comparing them with chance selections. The four choices, taken two at a time, can mathematically result in 12 permutations. Hence, each pair of choices has a chance level of one-twelfth. Therefore, only two subjects would be expected to select the same choices by chance (1/11 X 26 = 2.2). Consequently, selection of the same pair of choices by 15 subjects is seven times greater than chance. All the other choices were close to, at, or below the level of chance.

Reaction Scales

The responses of the fifteen Univesity of California subjects who had the same combination of most vs least preferred choices were used in the analysis of the Reaction Scales. The background of this group

does not appear to be different from those who had other combinations of choices. We analyzed the items on the Reaction Scales by doing t-tests for correlated means on each item. The items with their most vs least preferred Likert reaction scores are shown in Table 1. The 25 items which significantly differentiated the preferences are shown with an asterisk next to the t-value.

Although responses on the Reaction Scale for the most preferred report led to more favorable reactions than the least preferred report, the subjects agreed with only 9 of the 32 items for their most preferred report. In general, they found that repeated testing, content-bound, is helpful to parents and teachers (Reaction Scale items 1, 2, 3), informs them on student improvement (4), indicates how the student compares with other students (9), reveals that the student can read more difficult materials this year than last year (23), and informs the teacher on progress made in reading since last year (25). They also agreed that such information for the entire class would be useful in evaluating the teacher's or school's reading program (27). They disagreed that this kind of report would not tell teachers any more than they would already know as the child's teacher (22) and they also disagreed that tests provide little useful information (28).

By inference the subjects agreed that standardized tests provide useful information to teachers when reported for repeated testing, content-bound, but they were uncertain about the value of peer comparisons only and about content-free reporting of results (10). They

Table 1

Significance of Differences on Reaction Scale Items Between Most and Least Preferred Test Reports

Item	Mean Scores on Likert Type Scale (1=Most Agree; 5=Most Disagree) Rank Order of Test Report: Most Preferred	Rank Order of Test Report: Least Preferred	Statistics: t[b]	Statistics: Sig.
1. This information would be helpful in determining what instruction to provide for the student.	2.1[a]	4.3	6.1	**[c]
2. This information would be helpful in explaining this student's reading performance to parents.	1.8	4.1	6.5	**
3. This kind of report would allow parents to determine whether their child had improved compared to his past performance.	1.8	4.2	7.8	**
4. This student is improving in reading.	1.9	2.7	3.4	**
5. Tests tell whether teachers are doing a good job.	3.4	4.2	2.8	**
6. This information tells what the student is able to do.	3.3	4.1	3.4	**
7. This information indicates what kinds of questions on reading materials that the student can answer.	2.6	4.3	7.4	**
8. This type of information would help in planning for special instruction.	2.8	4.3	3.7	**
9. This information tells how the student compares with other students.	1.9	2.3	1.8	*[d]
10. Standardized tests do not provide useful information for classroom teachers.	3.7	3.7	-1.0	NS[c]
11. This information would help parents select materials appropriate for their child.	3.1	4.6	5.5	**

Table 1 (Cont.)

Mean Scores on Likert Type Scale
(1=Most Agree; 5=Most Disagree)

Item	Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t[b]	Sig.
12. This information tells me what kinds of materials the student can successfully read and answer questions on.	2.3	4.6	6.3	**
13. On the basis of this information, I could advise this student's next teacher about his instructional needs.	2.6	4.0	4.1	**
14. If parents were shown this kind of report, they could determine how their child compares to his classmates.	2.9	4.0	3.5	**
15. This information indicates what degree of mastery the student has attained on reading materials over the past year.	3.2	4.3	3.4	**
16. Percentile scores tell the percentage of items the student got right.	4.1	4.1	0.6	NS
17. This kind of information tells a teacher what kinds of materials to use for teaching a student.	3.4	4.6	3.7	**
18. This kind of information would inform the principal about how students are doing in my class.	3.0	4.0	2.6	*
19. Test results such as this will help the teacher group students for instruction.	2.6	4.7	3.7	**
20. This kind of information would just cause trouble with parents who would misunderstand it.	3.5	4.2	-1.5	NS
21. Tests tell you whether a student is doing the best he can.	4.3	4.5	1.1	NS

Table 1 (Cont.)

Mean Scores on Likert Type Scale
(1=Most Agree; 5=Most Disagree)

Item	Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t^b	Sig.
22. This kind of test report does not tell teachers any more than they would already know as a child's teacher.	3.8	2.5	-2.1	*
23. This information tells whether the student can read more difficult materials than last year.	2.0	3.6	4.3	**
24. This information indicates what kind of job the school is doing in teaching this student to read.	3.3	4.1	3.4	**
25. This information would help me inform the student about how much progress he has made in reading since last year.	2.2	3.8	5.9	**
26. This information indicates that as the student's second-grade teacher, I was effective in helping the student learn what he had not yet learned in the previous grade.	3.0	4.9	1.0	NS
27. Information of this type for all the students in my class would be useful in evaluating my reading program.	2.3	4.9	4.3	**
28. This information indicates that tests provide little useful information; they most likely take up teacher and pupil time.	4.3	3.1	-3.9	**
29. This kind of information would help a teacher in determining why a student is not able to read better.	3.5	4.3	3.2	**
30. Test scores tell you the quality of the school the student is in.	3.8	4.1	0.9	NS

Mean Scores on Likert Type Scale
(1=Most Agree; 5=Most Disagree)

Item	Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t[b]	Sig.
31. This information tells the student what kinds of materials he can read successfully.	2.6	4.1	5.0	**
32. This information indicates that the student can read materials at about the same difficulty level as last year.	3.7	3.5	0.2	NS

[a]Underline represents agreement with the statement.
[b]t at the .95 level (df 14) = 1.76; t at the .99 level (df 14) = 3.01
[c]Sig. at .01 level
[d]Sig. at .05 level
[e]NS = not significant

tended to disagree that repeated testing, with content-bound information would just cause trouble with parents who would misunderstand it. However, they were uncertain about the effect that peer comparisons, content-free, would have on parents (20). They were also uncertain whether the repeated testing, content-bound, indicated teachers were effective in helping students learn what they had not learned in the previous grade, but they were certain that the peer comparison, content-free did not provide them with this information (26).

The subjects know that percentile scores do not tell the percentage of items the students do right (16). They also recognize that tests do not tell whether students are doing the best they can nor do they inform students on what kinds of material they can read successfully.

This study was replicated on 18 educators enrolled at the University of Maryland, College Park. Twelve of them also had the same choice combinations as those in the first study. This number is eight times greater than chance. The other choices were again near, at, or below chance. The performance of the 12 subjects on the Likert-type Reaction Scale were almost identical to those on the first study.

Thus, the professional educators enrolled in courses at the University of California and the University of Maryland agreed that repeated testing with content-bound was preferable to the peer comparison, content-free test report for a variety of reasons. Since schools tend to report test results only in a content-free form, mostly emphasizing only peer group comparisons, which the professionals in this study reported is less preferable, the question is why

do schools not switch? At this point, we can only report on what we know and can only speculate on what we do not know.

Discussion

We know that the National Education Assessment Program provides content-bound information whenever it reports test results, usually in the form of percent of an age group that passes a particular item. We also know that reading specialists and clinicians routinely test prior to and retest after intervening instruction to their students. In short, retesting and reporting results with content-bound information does not involve new ideas, new tests, nor new technology. We also now know that professional educators would prefer such testing and test reporting, even though it involves adding repeated testing to the testing schedule.

Although some school districts do administer the same tests at the beginning and ending of the school year, such as the California Tests of Basic Skills, they still report only peer group comparisons when they use the test's fall and spring norms for interpreting and reporting test results. We advocate and used in our test reports on repeated testing the same norms for the tests given one year apart. Consequently we were able to reveal self-comparison gains.

Test reports now usually reveal only peer group comparisons in the form of percentiles or stanines based on peer group norms. These scores result frequently in a more or less horizontal line which laymen incorrectly perceive as indicating no gain. For a whole class, grade equivalent scores would also provide the perception of gains.

do schools not switch? At this point, we can only report on what we know and can only speculate on what we do not know.

Discussion

We know that the National Education Assessment Program provides content-bound information whenever it reports test results, usually in the form of percent of an age group that passes a particular item. We also know that reading specialists and clinicians routinely test prior to and retest after intervening instruction to their students. In short, retesting and reporting results with content-bound information does not involve new ideas, new tests, nor new technology. We also now know that professional educators would prefer such testing and test reporting, even though it involves adding repeated testing to the testing schedule.

Although some school districts do administer the same tests at the beginning and ending of the school year, such as the California Tests of Basic Skills, they still report only peer group comparisons when they use the test's fall and spring norms for interpreting and reporting test results. We advocate and used in our test reports on repeated testing the same norms for the tests given one year apart. Consequently we were able to reveal self-comparison gains.

Test reports now usually reveal only peer group comparisons in the form of percentiles or stanines based on peer group norms. These scores result frequently in a more or less horizontal line which laymen incorrectly perceive as indicating no gain. For a whole class, grade equivalent scores would also provide the perception of gains.

However, they would not differentiate acquisition from ability to gain information from text unless they were used with repeated testing. Moreover, grade equivalent scores have fallen into disuse.

Why do schools persist in testing for and reporting only peer group comparisons, content-free? In the absence of evidence, we can only speculate.

Speculation on Testing and Reporting Test Results

The most preferred test report, repeated testing, content-bound, also provides the most information. We can only conjecture on why schools do not use this type of testing, particularly in the primary grades when students are in the acquisition stage of development for basic skills. Nor do we know why schools do not provide content-bound information in their reports to pupils, parents, professionals, and the press. We recognize that more costs are involved, but they are not prohibitive. Repeated testing, of course, involves some additional cost, but the content-bound feature can be readily programmed into computers at some relatively small initial cost and printed out in reports with little additional cost.

Copyright on tests might be a problem. However, the Truth in Testing law has mandated release of test item information. In content-bound reporting only a few items would be released to any one student. However, if copyright or test control still constitutes a problem, then comparable items from previous editions of a test might be used with the appropriate qualification that the item is similar to the one the student had taken.

Possibly teachers might object to repeated testing because it entails more testing. But, the teachers who obtained the repeated test data on their classes for this study were delighted because the results were so gratifying to them. Hence, we think that teachers would welcome more testing when the results are likely to be favorable.

However, when results are returned to teachers one to three months after students have been tested, as often happens with machine-scored tests, they are likely to be perceived as irrelevant. The remedy for this problem is to have the teachers or their aides score tests and record the scores before sending the tests off for machine scoring.

Perhaps schools are reluctant to provide more than peer group comparisons, content-free, because of lack of knowledge of other ways of reporting, lack of pressure to report in more meaningful ways, or perhaps because they have qualms about providing more meaningful test results. We do not know the reason(s), but we intend to find out.

We also want to investigate some other questions. We want to know how repeated testing, content-bound, affects instructional decisions? How does it affect students' self attributions, self-esteem and locus of control if they perceive they are increasing rapidly in acquisition of basic skills, even through their peer-group comparisons are relatively low? How would parents react to the differentiated information with content-bound? Would newspaper and public reactions to schools be different if test results in basic skills were differentiated and reported with content-bound information for the average student? In general, we know that acquisition of basic skills

is very rapid in the primary grades. However, this knowledge is not known to the public. If it were, it might change their attitudes about school, particularly about instruction in the primary grades, and teachers' attitudes about the effectiveness of their instruction. Although the questions we have posed are not complex, we think that if answers to them lead the schools to adopt and report on repeated testing with content-bound information, for all students in the primary grades and for particular students in the intermediate grades who are still in the primary stage of reading acquisition, it will profoundly affect pupils, parents, professionals, and the press' perceptions of student and school accomplishments in development of the basic skills. If schools also reported peer group comparison data, content-bound for all students in grades 7-12, the results would be more meaningful; all concerned with the test results would then know what the students who took the tests know and can do.

Policy Implications

If schools continue with reporting of peer group comparisons, only, content free, then the reaction of teachers and administrators is likely to be negative. For a more positive reaction, schools can do differentiated testing in the elementary grades, especially over grades 1-6 when students are learning how to read, perhaps testing at the beginning and ending of each grade. Throughout the grades, they should always report test results and interpretations with content-bound information. That is, in reporting test results on reading comprehension the school should indicate the most difficult paragraph a student can read successfully on the test. When average scores are reported for a total class or grade, then the paragraph

that represents the average level attained on the test should be reported. If schools do not voluntarily adopt this practice after an educational campaign has been conducted to persuade them to do so, then the legislatures should mandate the requirement in their Educational Codes.

Summary and Conclusions

A group of 26 educators enrolled in a course in an administration program ranked from most to least preferable four types of reading achievement reports for the second grade student. The reports represented either repeated testing with peer group comparisons or peer group comparisons alone. Both types of reports were presented with and without content type information. Fifteen of the 26 educators ranked the test report based on repeated testing with content-bound information (most difficult paragraph correct on each test) as most preferable; they ranked the usual way of reporting test results in peer comparison only, content free form, as least preferable. The selection by 15 subjects of the same combination of most vs least preferable report was seven times greater than would occur by chance alone. The degree of agreement to disagreement by the 15 subjects to statements on a Likert type Reaction Scale indicated why they most preferred one type and least preferred the other. They agreed that their preferred type was useful to them and to parents and would help them evaluate students and the instructional program.

Although educators may have an antipathy to standardized achievement tests, the results of this study indicate that their antipathy is not directed at the tests themselves, but at the way in which tests

are used and the way in which they are reported. They tend to accept them when used in repeated testing and reported with content bound to the results, but they tend to reject them when used only for peer group comparisons and reported content-free. The implication is to switch to repeated testing, content-bound. It would be more helpful and useful to teachers, reveal to parents and pupils that progress in learning to read has occurred, and probably educate the public that schools succeed in teaching most students how to read and how to increase their ability to gain information from text. However, the public will also have to learn that schools cannot eliminate the range of individual differences in ability to gain information from text; in fact, successful schools help increase this range. Perhaps differentiated testing and content-bound reporting will enable the public to make the distinction that almost all students can successfully attain a criterion level in learning how to read, but all students cannot achieve the same criterion level in ability to learn from text. Indeed students can become alike in learning how to read, but they become more and more dissimilar in their ability to learn from text as they progress through school. Whether schools will be able to accomplish this differentiated understanding in pupils, parents, professionals, policymakers, and the press, remains to be seen. We do know that the sample used in this study least preferred peer group comparisons alone, the current way of reporting test results, but most preferred repeated testing with content-bound.

Footnotes

[1]This study was supported by funds from a California Policy Seminar Grant, administered by the Institute of Governmental Studies (John Cummins, Director). The grant was awarded to Harry Singer, John McNeil, Robert Ruddell, and Merl Wittrock, University of California (Riverside, Los Angeles, Berkeley, and Los Angeles, respectively) to study "Achievement Testing in Basic Skills."

[2]In a preliminary study, we had students rank order six categories: three types of testing (peer comparison, repeated testing, and both) with content-bound and with content-free. Subjects overwhelmingly preferred the combination testing, content-bound. Consequently in this study, we reduced the categories from 6 to 4 by eliminating the repeated testing alone report. We kept the peer group comparison alone, content-free, type of report because schools typically use it when reporting test results.

References

Bloom, B. S. Mastery learning and its implications for curriculum development. In E. W. Eisner (ed.), Confronting Curriculum Reform. Boston: Little, Brown, 1971.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Cook, W. W. The functions of measurement in the facilitation of learning. In E. F. Lindquist (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1951.

Green, J. Reporting test results in New York City. National Elementary Principal, 1975, 54(6), 91-92.

Goslin, D. A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Hilgard, E. R., & Bower, G. H. Theories of learning. New York: Appleton-Century-Crofts, 1966.

Lerner, B. The war on testing: David, Goliath, and Gallup. The Public Interest, 1980 (Summer), 121.

Madaus, G. F. Reactions to the Pittsburgh papers. Phi Delta Kappan, 1981, 62(9), 634-636.

National Education Research. Teacher opinion poll: Group standardized tests. Today's Education, 1978, 67(4), 20.

Resnick, D. P. Testing in America: A supportive environment: Phi Delta Kappan, 1981, 62(9), 625-628.

Rothkopf, E. Adjunct Aids and the control of mathemagenic activities during reading. Invitational Conference on Reading Expository Material. University of Wisconsin, Madison: Research and Development Center, November 10-11, 1980.

Salmon-Cox, L. Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 1981, 62(9), 631-634.

Singer, H. IQ is and is not related to reading. In S. Wanat (Ed.), Issues in Evaluating Reading. Arlington, Virginia: Center for Applied Linguistics, 1977a.

Singer, H. Measurement of early reading ability. Contemporary Education, 1977b, 48(3), 145-150.

Singer, H., McNeil, J., Ruddell, R. B., & Wittrock, M. Testing achievement in basic skills. (Final report.) University of California: California Policy Seminar, Institute of Governmental Relations, Berkeley, California (in preparation; expected date of completion: August 1982.)

Sproul, L., & Zubrow, D. Standardized testing from the administration perspective. Phi Delta Kappan, 1981, 62(9), 628-631.

Stetz, F. P., & Beck, M. D. Teachers oppose halting all standardized testing. Phi Delta Kappan, 1980, 61(9), 648.

Takeuchi, R. T. Attitudes of elementary teachers toward testing: Use and abuse of standardized tests in California, 1976-1977. Dissertation Abstracts, 1978, 38(8-A), 4500.

Tyler, R. W., & White, S. H. (Editors). Testing, teaching and learning. (Report of a Conference on Research in Testing, August 17-26, 1979.) Washington, D. C.: U.S. Department of Health, Education, and Welfare, 1979.

D143-Revised 9/82

Attitudes Towards Testing and Test Results in Achievement of Basic Skills

Harry Singer
University of California, Riverside

and

Mariam Jean Dreher
University of Maryland,
College Park

September 1982

Attitudes Towards Testing and Test Results in Achievement of Basic Skills

Abstract

Four ways of testing and reporting test results were ranked ordered from most to least preferred by a group of educators enrolled in an administration course at the University of California, Riverside. They also completed Likert-type Reaction Scales for each of their accepted and repeated choices. The four choices were based on two tests: (1) standardized reading test administered in grade 1.8 and readministered in grade 2.8; both test results were interpreted according to grade 1.8 norms; the graphed results was called self-comparison. (2) A second standardized test, normed referenced for and administered in grade 2.8; a graph of the 1.8 and 2.8; a graph of the 1.8 and 2.8 results for the two different tests was labeled peer-comparison. Another factor was content-free vs. content-bound reporting of test results. The content-free graph of the results depicted only numerical information; the content-bound graph contained numerical information plus the most difficult paragraph and questions a student could successfully read on each test. The most preferred graph of test results contained self- and peer-comparisons with content-bound. The least preferred graph consisted of peer-comparison only, content-free. Although responses to the 32 Likert Scale items revealed statistically significant differences between the two

choices, the educators expressed agreement with only 9 of these items. These items indicated they agreed that the combination of self and peer comparison testing and test results, content-bound is helpful to parents and teachers, informs them on student improvement, indicates how a student compares with others, reveals the student could read more difficult materials after another year of instruction, informs the teacher on student progress, provides teachers with information they did not already know, and thus the tests and test results provide them with useful information. The study was completely replicated on a similar group of educators at the University of Maryland. Hence, the ambivalent and somewhat negative attitudes of educators towards tests and test results can be reliably shifted in the direction of more favorable responses by tests and test results that are more diagnostic and content-bound. Consequently, the results call for a change in current practices in testing and reporting of test results in basic skills.

Dr. Harry Singer
School of Education
University of California
Riverside, CA 92521
714-787-5228 (office)
714-686-9195 (home)

Dr. Mariam Jean Dreher
Reading Center
University of Maryland
College Park, MD 20742

Attitudes Towards Testing and Test Results in Achievement of Basic Skills

Harry Singer[1,2,3]
University of California,
Riverside

Mariam Jean Dreher
University of Maryland
College Park

The use of standardized tests in the United States has increased over the past 30 years.[4] This receptive atmosphere for standardized tests stems historically from American ideals for objective selection of talented students regardless of social origins, desire for efficient use of resources and rational placement and grouping of students so as to reduce the failure rate, and national standards for evaluating school achievement and thereby limiting autonomy of local school districts (Resnick, 1981).

However, school personnel think that standardized tests are important, but for some one else's purpose! For example, when central office administrators in 18 school systems (public, private, and parochial) in Western Pennsylvania were surveyed, the public school administrators reported that principals and teachers were the school personnel that benefit the most from tests. But, principals in private schools stated that parents, students, and communities have the greater need for test results than do school personnel (Sproul & Zukow, 1981).

However, we do know that despite the criticism of tests from professionals, civil rights workers, and consumer advocates, parents, business people, and legislators want more testing (Madaus, 1981). In a national Gallup poll conducted in 1979 parents overwhelmingly responded that standardized test scores were "useful" or "somewhat useful" (Lerner, 1980). Although some organizations, such as the Council for Exceptional Children and the National Education Association have called for a moratorium on testing, a 1978 survey found that only 16 percent of 3500 teachers in a national sample agreed with this position (Stetz & Beck, 1980). However, California teachers have an ambivalent attitude towards standardized tests. Ninety-two percent did not favor their use, but 84 percent reported they would be unable to teach effectively without them (Takeuchi, 1978). Another national poll, conducted by the National Education Association, found that 84 percent of the teachers in the survey felt that standardized test scores could be misleading and biased but 90

percent nevertheless think that standardized tests are appropriate when not used in isolation (NEA, 1978).

Apparently teachers have more confidence in their own judgments about students than they have in test results. For example, 68 teachers in a suburban school district near Pittsburgh and a large urban school district noted that standardized tests, such as the Metropolitan, Stanford, and California Test of Basic Skills were useful in supplementing their information on students, but placed more credence in their own observation of children's performance in the classroom than on standardized test scores, particularly when test scores were lower than expected. When they were higher, teachers reexamined their assumptions about students. The teachers also stated a need for tests that are diagnostic, closely match the curriculum and instruction, and yield results quickly (Salmon-Cox, 1981). Administrators and teachers in Ireland where testing was introduced for experimental purposes arrived at similar opinions after only four years of experience with standardized tests (Madaus, 1981). However, in neither situation did school personnel indicate the kind of diagnostic information they wanted.

The results of all of the studies cited have been based on current test practices which usually involve the reporting of only numerical test data such as percentiles and stanines, particularly when test results appear in the newspapers (Green, 1975). Although some experimentation in testing has been conducted since Goslin's (1967) well-known study, such as Madaus's (1981) and those reported in

a recent conference on testing (Tyler & White, 1979), a search of the current research literature failed to disclose any experimentation on ways of testing and reporting test results. Perhaps the controversy between professionals and the public can at least be reduced and perhaps the ambivalence of educators towards tests can be more or less resolved by testing and reporting of test results that more closely satisfy their request for tests that are more diagnostic and sensitive to their curriculum and instruction (Singer, 1977a,b; Singer, McNeil, Ruddell, and Wittrock, 1982). The problem is to determine what diagnostic information would be satisfactory to school personnel and whether it would shift their attitudes toward a favorable response to testing and test results.

In the first part of our study we determined the kind of testing and test results from among four choices that a sample of educators prefer (Singer and Dreher, 1982). The choices are shown in Figures 1-4. Figure 1 consists of the current way of testing and reporting test results: grade level standardized tests administered in grades 1.8 and 2.8 with a report of percentile ranks only, as shown by the dashed line in Figure 1. We call Figure 1 "peer comparison, numerical data only." Figure 2 includes peer comparison data plus the results of a retesting at the end of grade 2.8 on the same standardized test that had been administered previously in grade 1.8. The results of the initial and the second testing are shown by the solid line in Figure 1. We call this type of testing and reporting, "peer and self-comparison, numerical data only." Figures 3 and 4 consist of

presenting the information in Figures 1 and 2, respectively, with the most difficult paragraphs that the students could read successfully. These paragraphs are shown on the sides of Figure 3 and 4. We call this addition to the test results, "content bound information." Two groups of educators, consisting of teachers and administrators at the elementary, junior, and senior high school levels overwhelmingly preferred Figure 4: self-comparison and peer comparison with content-bound information.

The next question is what effect the most preferred choice vs least preferred choice (peer comparison only without content bound information) would have on the attitudes of educators towards testing and test results. Would a change in testing and reporting of test results, as indicated by their preferred choice, significantly change their attitudes towards testing and test results? To answer this question, we constructed a Reaction Scale towards Testing and Test Results and administered them in a balanced experimental design.

Experimental Design

After the experienced school personnel during the first part of our study (Singer and Dreher, 1982) had indicated their most and least preferred test results, they filled out a Likert-type Reaction Scale consisting of 32 items, once for their most preferred and again for their least preferred category of test result. The scales were done in counter-balanced order: Half the subjects first reacted to their

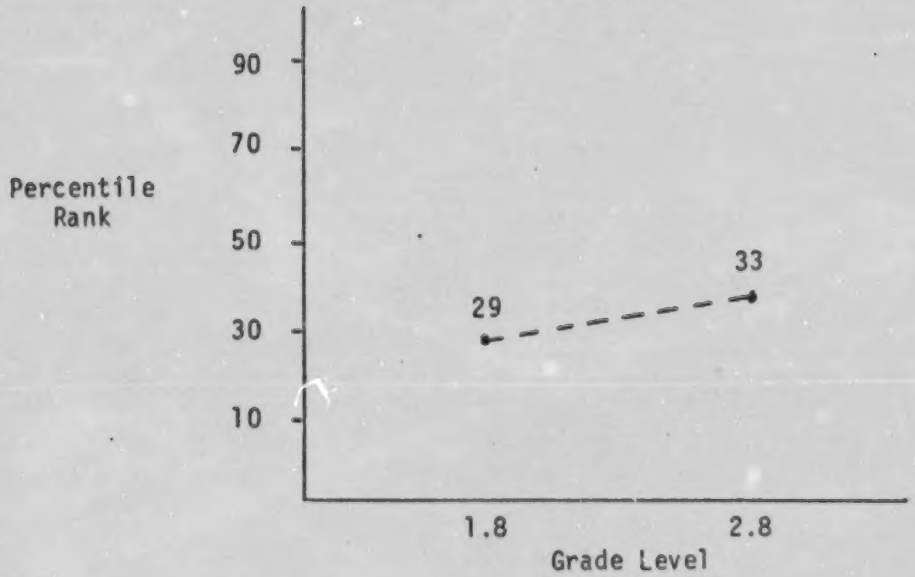


Figure 1. Current way of reporting test results: peer-comparison, numerical data only. The example shown above indicates the percentile ranks for a student tested at the end of grades 1.8 and 2.8 on a standardized test, norm referenced for each grade.

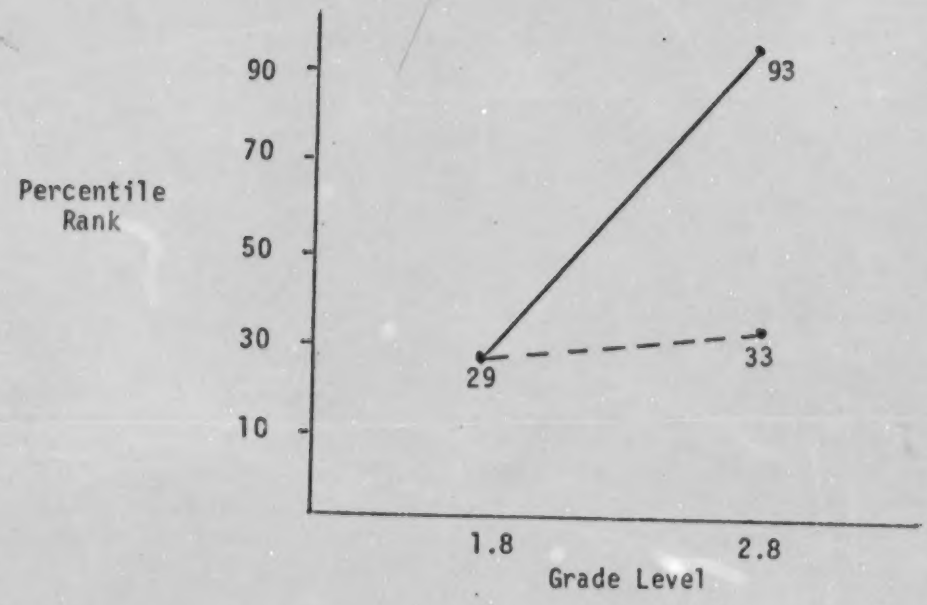
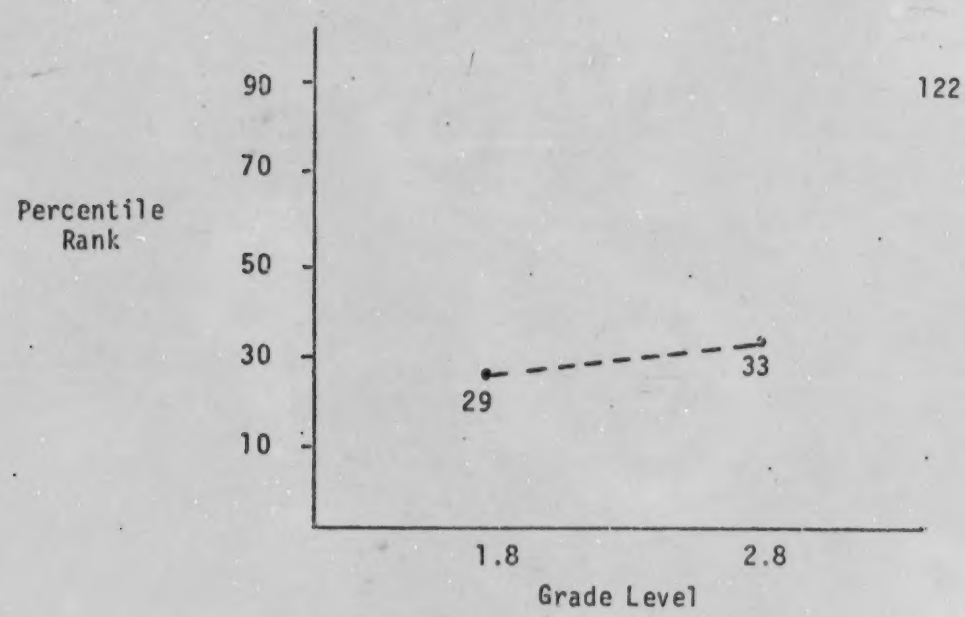


Figure 2. New way of testing and reporting test results: peer-comparison (dashed line) and self-comparison (solid line), numerical data only. The example above shows the results for three tests. The dashed line indicates the same peer-comparison data shown in Figure 1. The solid line represents self-comparison data, the results of repeated testing: testing the same student at the end of grades 1.8 and 2.8 on the same test and using the same grade 1.8 norms to interpret the results.



Content Bound Information

This is the hardest test passage the student was able to read correctly on the first-grade test at the end of first grade. Student is to fill in the circle under the picture that represents the sentence

This is the hardest test passage the student could reade correctly on the second-grade test at the end of second grade.

It was Ann's birthday. Mr. Parks had a surprise for Ann. "Here is the box of candy," he said. "Thank you, Father," said Ann, "I will put it on my toy box. I must go to school now, Betty Smith is waiting for me."

Frisky barked when he saw Ann go away. Then he ran to the toy box and got the candy. Frisky had a good time.

What did Frisky do?
- ran away
- ate the candy
- went to school

What was Father's last name?
- Brown
- Frisky
- Smith

Figure 3. Reporting on current way of testing, tests norm referenced for each grade, but with content-bound. This type of report is called peer-comparison, content-bound.

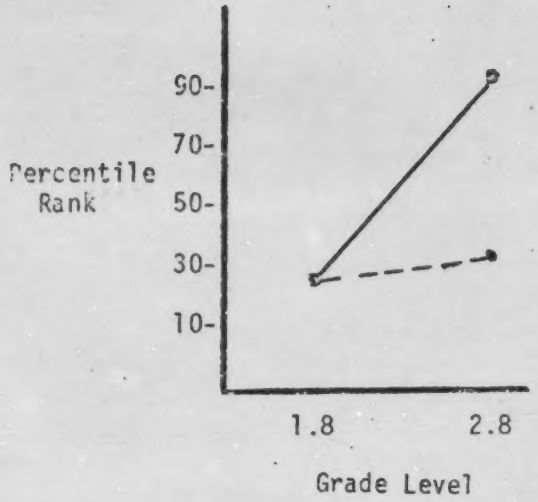
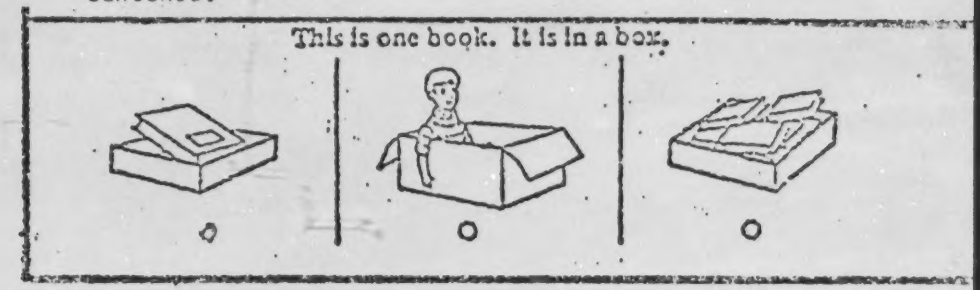


Figure 4. Reporting on new way of testing: Peer-Comparison and Self-Comparison results with Content-Bound.

This is the hardest test passage the student was able to read correctly on the first-grade test at the end of first grade. Student is to fill in the circle under the picture that represents the sentence.



This is the hardest test passage the student read correctly on the first-grade test at the end of second grade. (Note: He got all the passages correct.) Student is to fill in the circle under the picture that represents the sentence.

This is the hardest test passage the student could read correctly on the second-grade test at the end of second grade.

It was Ann's birthday. Mr. Parks had a surprise for Ann. "Here is the box of candy," he said. "Thank you, Father," said Ann, "I will put it on my toy box. I must go to school now, Betty Smith is waiting for me."

Frisky barked when he saw Ann go away. Then he ran to the toy box and got the candy. Frisky had a good time.

What did Frisky do?
- O ran away
- O ate the candy
- O went to school

What was Father's last name?
- O Brown
- O Frisky
- O Smith

most preferred, then to their least preferred choice. The other half had the reverse order. Each person indicated on each Reaction Scale his or her most and least preferred type of test result.

Subjects

A total of 26 subjects participated in the University of California study. They were students enrolled in an administration credential course who had an average of 5.1 years teaching experience; eight taught at the elementary, four at the junior high, and five at the high school level. Five were administrators at the elementary or high school levels and two had some administrative position. The modal number of courses in tests and measurements taken by the group was one. Two subjects did not complete the information inventory.

The study was repeated at the University of Maryland on 18 students enrolled in a graduate course in research in reading. Thirteen of the students had seven or more years of teaching experience; the others all had from four to six years of teaching experience. Nine were currently teaching at the preschool through elementary level, two at the junior high, and one at the high school level. Two were principals of elementary schools and one had another administrative position. Three subjects neglected to indicate their current positions but were known to be experienced teachers. The modal number of courses in tests and measurements taken by the group was two. Both groups were, of course, given the same directions.

Directions

Keeping your most preferred report in mind, read and record your reaction to each statement on the first Reaction

Scale. Then, record your reaction for your least preferred report on the second Reaction Scale. (Most preferred and least preferred were, of course, reversed for half of the group.)

Analysis of Results

The responses of the fifteen University of California subjects who had the same combination of most vs least preferred choices were used in the analysis of the Reaction Scales. The background of this group does not appear to be different from those who had other combinations of choices. We analyzed the items on the Reaction Scales by doing t-tests for correlated means on each item. The items with their most vs least preferred Likert reaction scores are shown in Table 1. The 25 items which significantly differentiated the preferences are shown with an asterisk next to the t-value.

Although responses on the Reaction Scale for the most preferred report led to more favorable reactions than the least preferred report, the subjects agreed with only 9 of the 32 items for their most preferred report. In general, they found that the combination of self-comparison and peer-comparison content-bound (Figure 4), is helpful to parents and teachers (Reaction Scale items 1, 2, 3), informs them on student improvement (4), indicates how the student compares with other students (9), reveals that the student

Table 1

Significance of Differences on Reaction Scale Items Between Most and Least Preferred Test Reports

Reaction Scale towards Testing and Test Results Item	Mean Scores on Likert Type Scale (1=Most Agree; 5=Most Disagree) Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t[b]	Sig.
1. This information would be helpful in determining what instruction to provide for the student.	2.1[a]	4.3	6.1	**[c]
2. This information would be helpful in explaining this student's reading performance to parents.	1.8	4.1	6.5	**
3. This kind of report would allow parents to determine whether their child had improved compared to his past performance.	1.8	4.2	7.8	**
4. This student is improving in reading.	1.9	2.7	3.4	**
5. Tests tell whether teachers are doing a good job.	3.4	4.2	2.8	**
6. This information tells what the student is able to do.	3.3	4.1	3.4	**
7. This information indicates what kinds of questions on reading materials that the student can answer.	2.6	4.3	7.4	**
8. This type of information would help in planning for special instruction.	2.8	4.3	3.7	**
9. This information tells how the student compares with other students.	1.9	2.3	1.8	*[d]
10. Standardized tests do not provide useful information for classroom teachers.	3.7	3.7	-1.0	NS[e]
11. This information would help parents select materials appropriate for their child.	3.1	4.6	5.5	**

Mean Scores on Likert Type Scale
(1=Most Agree; 5=Most Disagree)

Item	Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t^b	Sig.
12. This information tells me what kinds of materials the student can successfully read and answer questions on.	2.3	4.6	6.3	**
13. On the basis of this information, I could advise this student's next teacher about his instructional needs.	2.6	4.0	4.1	**
14. If parents were shown this kind of report, they could determine how their child compares to his classmates.	2.9	4.0	3.5	**
15. This information indicates what degree of mastery the student has attained on reading materials over the past year.	3.2	4.3	3.4	**
16. Percentile scores tell the percentage of items the student got right.	4.1	4.1	0.6	NS
17. This kind of information tells a teacher what kinds of materials to use for teaching a student.	3.4	4.6	3.7	**
18. This kind of information would inform the principal about how students are doing in my class.	3.0	4.0	2.6	*
19. Test results such as this will help the teacher group students for instruction.	2.6	4.7	3.7	**
20. This kind of information would just cause trouble with parents who would misunderstand it.	3.5	4.2	-1.5	NS
21. Tests tell you whether a student is doing the best he can.	4.3	4.5	1.1	NS

Item	Mean Scores on Likert Type Scale (1=Most Agree; 5=Most Disagree) Rank Order of Test Report		Statistic[a]	
	Most Preferred	Least Preferred	t[b]	Sig.
22. This kind of test report does not tell teachers any more than they would already know as a child's teacher.	3.8	2.5	-2.1	*
23. This information tells whether the student can read more difficult materials than last year.	2.0	3.6	4.3	**
24. This information indicates what kind of job the school is doing in teaching this student to read.	3.3	4.1	3.4	**
25. This information would help me inform the student about how much progress he has made in reading since last year.	2.2	3.8	5.9	**
26. This information indicates that as the student's second-grade teacher, I was effective in helping the student learn what he had not yet learned in the previous grade.	3.0	4.9	1.0	NS
27. Information of this type for all the students in my class would be useful in evaluating my reading program.	2.3	4.9	4.3	**
28. This information indicates that tests provide little useful information; they most likely take up teacher and pupil time.	4.3	3.1	-3.9	**
29. This kind of information would help a teacher in determining why a student is not able to read better.	3.5	4.3	3.2	**
30. Test scores tell you the quality of the school the student is in.	3.8	4.1	0.9	NS

Table 1 (Cont.)

Mean Scores on Likert Type Scale
(1=Most Agree; 5=Most Disagree)

Item	Rank Order of Test Report		Statistics	
	Most Preferred	Least Preferred	t[b]	Sig.
31. This information tells the student what kinds of materials he can read successfully.	2.6	4.1	5.0	**
32. This information indicates that the student can read materials at about the same difficulty level as last year.	3.7	3.5	0.2	NS

[a]Underline represents agreement with the statement.
[b]t at the .95 level (df 14) = 1.76; t at the .99 level (df 14) = 3.01
[c]Sig. at .01 level
[d]Sig. at .05 level
[e]NS = not significant

can read more difficult materials this year than last year (23), and informs the teacher on progress made in reading since last year (25), and tells the student what kind of materials he or she can read successfully (31). They also agreed that such information for the entire class would be useful in evaluating the teacher's or school's reading program (27). They disagreed that this kind of report would not tell teachers any more than they would already know as the child's teacher (22) and they also disagreed that tests provide little useful information (28).

By inference the subjects agreed that standardized tests provide useful information to teachers when reported for repeated testing, content-bound, but they were uncertain about the value of peer comparisons only and about content-free reporting of results (10). They tended to disagree that repeated testing, with content-bound information would just cause trouble with parents who would misunderstand it. However, they were uncertain about the effect that peer comparisons, content-free, would have on parents (20). They were also uncertain whether the repeated testing, content-bound, indicated teachers were effective in helping students learn what they had not learned in the previous grade, but they were certain that the peer comparison, content-free did not provide them with this information (26).

The subjects were knowledgeable about testing and test results. They know that percentile scores do not tell the percentage of items the students do right (16). They also recognize that tests do not tell whether students are doing the best they can (21).

The results of this study are not only replicable but also robust. The data from the replication at the University of Maryland were almost identical to those obtained from the educators participating in the University of California study. This replication means that the results are not only statistically but also experimentally reliable. Moreover, an elementary school faculty expressed the same preferences from testing and reporting test results as did our two experimental samples. Hence, we are confident our results are also generalizeable.[5]

Discussion of Results

These results indicate that educators are not opposed to tests per se. They apparently are opposed to tests and test results that do not provide them with meaningful or additional information. When given testing and test results that do provide them with diagnostic information, their attitudes shifted, much to our pleasant surprise, from a negative to a positive response to testing and test results! We attribute this dramatic shift in attitudes towards testing and test results to the kind of test and test results shown in Figure 4. The diagnostic information shown in Figure 4 differentiates progress in learning how to read (self-comparison) from relative ability to comprehend or gain information from text (peer group comparison).

Although teachers in the primary grades teach students how to read and how to gain information from text, the usual way of testing and reporting test results confound these two phases of reading. Tests which are standardized and norm-referenced for the grade level in

which they are administered and interpreted according to these norms provide only peer-group comparisons. Under this condition, teachers do not obtain information on how their students are progressing in learning how to read. Teachers want and need both types of information. As we have demonstrated, it is possible to give them both types of information without having to construct any new tests. The only addition is repeated testing with the same group of students and with the same norms to interpret the test results. The resulting graph of the results of this self-comparison, as shown in Figures 2 and 4, depicts the diagnostic information in a perceptually compelling way that is easy to interpret. Moreover, when content bound information is added the results are made even more meaningful (Figures 3 and 4). The viewer then has the necessary information for knowing the most difficult level of tested content the student can successfully read.

In general, learning how to read has long been known to have a rapid growth curve in the primary grades (Buswell, 1922). But current testing practices have not revealed this information to educators and the public. However, it is relatively easy for schools to obtain and report such data. If schools did adopt the practice, then we predict that teachers would be delighted with the information as were the teachers in La Granada Elementary School in Alvord, California, where we collected the test data used in Figures 1-4. Moreover, we think that pupils, their parents, the public, and the press would soon conclude from such diagnostic information that almost all children make rapid progress in learning to read and that schools are quite successful in teaching reading.[6]

However, test consumers would also observe that peer group comparisons, relative ability to comprehend, does not change very much. We can use an analogy to better explain the difference between progress in learning to read (self-comparisons) with relative ability to comprehend (peer-group comparisons). The analogy will help explain why self-comparisons change dramatically while peer-comparisons reveal little or no change. In running a race, almost all students cover some distance from the starting line, which is analogous to progress in learning to read, the self-comparison data shown by the solid line in Figures 2 and 4. But also in the race, the relative positions of the students are not likely to change very much as they race around the track, which is analogous to ability to comprehend or gain information from text at any given point in a student's school career. This relative reading achievement is depicted as peer-group comparisons, the dashed line in Figures 1-4.

As a result of receiving reports on both learning to read and relative reading achievement, we think that neither educators nor the public are likely to misinterpret the two types of test results. Similar differentiation in test results could also be obtained for the other basic skills of arithmetic and writing. If schools were to differentiate testing and reporting of test results in arithmetic computation from arithmetic problem solving and in learning how to write from use of writing for self expression then we would probably observe similar changes in attitudes towards achievement in these basic skills too.

Conclusion

Although educators currently have an antipathy towards testing and test results, this study demonstrates that their antipathy is not directed at testing itself, but at the way in which testing is done and reported. They tend to reject testing and test results when used only for peer group comparison reported content-free. But they accept them when they are used in repeated testing that differentiates learning how to read from relative ability to learn from text and when the results are reported with content-bound information. In short, this type of testing and way of reporting test results fits teachers' desire for diagnostic information. In contrast to reactions to current test practice, differentiated testing and reporting of test results with content-bound information elicits a favorable attitude towards testing and reporting test results on achievement in basic skills.

Footnotes

[1]This study was supported by funds from a California Policy Seminar Grant, administered by the Institute of Governmental Studies (John Cummins, Director). The grant was awarded to Harry Singer, John McNeil, Robert Ruddell, and Merl Wittrock, University of California (Riverside, Los Angeles, Berkeley, and Los Angeles, respectively) to study "Achievement Testing in Basic Skills."

[2]In a preliminary study, we had students rank order six categories: three types of testing (peer comparison, repeated testing, and both) with content-bound and with content-free. Subjects overwhelmingly preferred the combination testing, content-bound. Consequently in this study, we reduced the categories from 6 to 4 by eliminating the repeated testing alone report. We kept the peer group comparison alone, content-free, type of report because schools typically use it when reporting test results.

[3]We wish to express our appreciation to Dr. Robert Hocker, acting superintendent, and to the principal, teachers, and students in La Granada Elementary School, Alvord, California for their cooperation in the testing phase of this study. We also are grateful to Mr. Ray Berry, Lecturer in Education, and to his students at the University of California, Riverside and to the students at the University of Maryland, College Park, for their participation in this study.

[4]Their sales have risen steadily from $7,000,000 in 1948 to $42,000,000 in 1976 (only $24,000,000 when corrected for inflation since 1967) (Resnick, 1981).

[5]We are grateful to Mrs. Deborah Flores, principal of Longfellow Elementary School, Riverside, California, and her faculty for participating in this study.

[6]Test consumers might also learn from the diagnostic information that almost all students in the normal range of intelligence, from about IQ 70 and up, are acquiring the basic skill of learning to read (Bloom, 1971, 1982). The only difference that IQ makes in this phase of reading is the rate at which students acquire this ability (Singer, 1977a,b) Under optimal school conditions, in which instruction is properly sequenced and students have adequate time for learning the tasks, the rate of learning to read is likely to be most rapid (Carroll, 1963; Barr, 1973-74). Concomitantly, as a group of students approaches mastery in learning how to read, the relationship would decrease between their IQ and reading achievement on the repeated administrations of the same reading test (Singer, 1977a).

However, for peer-group comparisons, IQ has an increasing relationship to reading achievement because as students progress through the grades IQ and reading tests become more alike and more similar in the mental processes they both tap (Singer, 1977a). Moreover, as a group of students progresses through school, individual differences in knowledge that students already have and can gain from text increase because the knowledge to be gained is infinite and individuals differ in rate of learning this knowledge.

References

Barr, R. Instructional pace differences and their effect on reading acquisition. Reading Research Quarterly, 1973-74, 9(4), 526-554.

Bloom, B. S. Mastery learning and its implications for curriculum development. In E. W. Eisner (ed.), Confronting Curriculum Reform. Boston: Little, Brown, 1971.

Bloom, B. Human characteristics and school learning. New York: McGraw-Hill, 1982.

Buswell, G. T. Fundamental reading habits; A study of their development. Supplementary Educational monographs, no. 21. Chicago: University of Chicago Press, 1922.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Green, J. Reporting test results in New York City. National Elementary Principal, 1975, 54(6), 91-92.

Goslin, D. A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Lerner, B. The war on testing: David, Goliath, and Gallup. The Public Interest, 1980 (Summer), 121.

Madaus, G. F. Reactions to the Pittsburgh papers. Phi Delta Kappan, 1981, 62(9), 634-636.

National Education Research. Teacher opinion poll: Group standardized tests. Today's Education, 1978, 67(4), 20.

Resnick, D. P. Testing in America: A supportive environment: Phi Delta Kappan, 1981, 62(9), 625-628.

Salmon-Cox, L. Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 1981, 62(9), 631-634.

Singer, H. IQ is and is not related to reading. In S. Wanat (Ed.), Issues in Evaluating Reading. Arlington, Virginia: Center for Applied Linguistics, 1977a.

Singer, H. Measurement of early reading ability. Contemporary Education, 1977b, 48(3), 145-150.

Singer, H., & Dreher, J. Preference for reporting of test results: Acquisition vs ability to gain information from text, content-bound and content-free. In J. Niles and L. Miller (Editors), Thirty-First Yearbook of the National Reading Conference, Rochester, N.Y.: The National Reading Conference, 1982 (in press).

Singer, H., McNeil, J., Ruddell, R. B., & Wittrock, M. Testing achievement in basic skills. (Final report.) University of California: California Policy Seminar, Institute of Governmental Relations, Berkeley, California (in preparation; expected date of completion: August 1982.)

Sproul, L., & Zubrow, D. Standardized testing from the administration perspective. Phi Delta Kappan, 1981, 62(9), 628-631.

Stetz, F. P., & Beck, M. D. Teachers oppose halting all standardized testing. Phi Delta Kappan, 1980, 61(9), 648.

Takeuchi, R. T. Attitudes of elementary teachers toward testing: Use and abuse of standardized tests in California, 1976-1977. Dissertation Abstracts, 1978, 38(8-A), 4500.

Tyler, R. W., & White, S. H. (Editors). Testing, teaching and learning. (Report of a Conference on Research in Testing, August 17-26, 1979.) Washington, D. C.: U.S. Department of Health, Education, and Welfare, 1979.

Classroom Testing

John McNeil
Education-Moore Hall
University of California
Los Angeles, CA 90024
(213-825-2700) (office)
(213-455-1296) (home)

Summarization Strategy for Improving Reading Comprehension

John McNeil[1]
University of California,
Los Angeles

Lisbeth Donant
Atascadero Public Schools

The focus of this study stems from a demonstration that the comprehension scores of sixth graders doubled when pupils generated sentences about paragraphs and inserted relevant paragraph headings in their sentences (Doctorow, Wittrock, and Marks, 1978). This demonstration led us to ask and seek an answer to this question: Will training in the rules of summary writing result in better summaries and greater reading comprehension?

Brown, Campione, and Day (1981) identified six basic rules essential to summarization. These rules are similar to the operations described by Kintsch and van Dijk (1978) as basic operations in comprehending and remembering prose. The six rules are (1) deleting unnecessary or trivial material; (2) deleting material that is important but redundant; (3) substituting a superordinate term for a list of items; (4) substituting a superordinate term for components of an action; (5) selecting a topic sentence; and (6) when there is no topic sentence, inventing one. Brown et al. (1981) found that the summaries produced by students from grades 5, 7, 10, and college show developmental differences. Students become increasingly adept at using superordination and selecting topic sentences. However, invention was rarely used by fifth graders and by college students on only half of the occasions when it was appropriate.

Day (1980) has been successful in teaching junior college students to apply the rules. The question remains, however, whether or not young children can be taught the rules and with what effect on reading comprehension?

Subjects

Twenty-three fifth grade pupils in a middle class school were randomly assigned to one of three groups: a summary rule training group, a summary writing group, and a non-instructional control group.

Procedures and Materials

Pretest

All students were given a pretest consisting of two separate passages. Questions posed about the passages required five inferences. There was no significant difference among pretest scores for the three groups.

Rule Training

Students in the rule training group were given group training on each of the six rules. The reading level of these materials was approximately two grade levels lower than the grade level of the students.

The basic instructional exercises for this training consisted of instructions for twelve passages, such as the following:

"Put a line through all the statements that are not important to the story."

> What kind of things make people feel nice? It's nice to get something. There couldn't be a better feeling. Or could there? Could it be better to give something? Think about it. Some people wish they could take a trip.

"Put a line through all the statements that repeat what has already been said."

Peter told the cat about the wolves that lived in the forest beyond a meadow. They were large and very, very mean. They were bigger than any dog and mean.

"What is this story about?"

Sometimes a sentence has marks like these around it (" ") and an end mark like this: (!). Sometimes there is an end mark like this (?). Can you tell how the speaker feels who says, "He found my bike!" and how another speaker feels when he says, "He found my bike?"

"What is this story about?"

Peg got out the paper. Bill found the paint. "Let's draw our street," said Andy. Soon there was a drawing.

"There's just one more thing," said Peg.

"See there's no sun in the drawing."

"Let me fix it."

"Put a line under the topic sentence for this paragraph."

Mirrors help you in many ways. They help you see yourself. The mirror in a car helps the driver see behind. A dentist might use a mirror to look at your teeth.

"Make up a topic sentence for this paragraph."

> We talk about the kind of seeds they plant. We talk about the kind lady. We laugh about the funny clown and we say it's funny when something strange happens. We can turn on our toes and we can turn on the light. Don't be mean - tell me what these words mean.

The first selection in each exercise was used to assess a pupil's status with respect to ability to apply the rule. If children could not apply the rule, the teacher modeled it and gave additional practice exercises. For example, if the pupils could not identify irrelevant parts, they were given such classification problems as, "table, chair, banana - Which one does not belong?" These problems were followed with activities by which pupils (a) listened to stories and raised their hands when hearing something that did not fit; (b) read paragraphs silently and drew lines under the unimportant parts. After these practice exercises, pupils were asked to apply the rule to the second selection.

Rules were taught in a successive order. Approximately 25 minutes daily was spent on each rule. An exception was in the amount of time given to Rule 5, finding a topic sentence. Students required less time for this rule because it was already familiar to them. Once all rules were taught in isolation, the teacher presented the training group with additional paragraphs taken from classroom materials and asked the pupils to apply each rule and then write a summary for the respective paragraphs. Summaries were not to exceed ten words and not to contain

unimportant details. Meanwhile, students in the two control groups carried out their daily classroom activities without special treatment.

Post-test

The post-test consisted of two selections representative of the reading passages found in the California Assessment Program for Assessing Reading Comprehension (California State Department of Education, 1980). The twelve corresponding questions called for the inferential skills for recognizing main idea, cause and effect, and drawing conclusions. Skills of noting details, recognizing pronoun referents, and sequence were also applicable.

Prior to administering the comprehension questions, the teacher asked those in the training group to apply the six rules to the selections and to write a summary for them.

Students in the summary only group were given a rule card listing the six rules and asked to apply them to the selections after which they were to write their summaries.

Students in the non-instructional control group were asked to read the selections and to be prepared for subsequent questioning.

The selections and summaries were collected and the written questions were given for individually written responses.

Results

Table I below shows the percentages of pre and post test scores correctly answered by pupils within the three conditions.

Table I

Mean Number and Percentage of Comprehension Questions Correctly Answered by Pupils Under Three Conditions (post-test)

	Training in Rules For Writing Summaries	Writing Summaries Without Training in Rules	Non-Instructional Control
Mean	10.2	7	9.2
Percentage	85%	58%	76%
	(N = 8)	(N = 7)	(N = 8)

Those receiving both training and opportunity to write summaries for the material excelled over both those who wrote summaries without rule training and those who did not receive training or opportunity to write summaries.

Analysis of the summaries written by those in the training and non-training groups showed that only one pupil out of the eight with training failed to write a summary that met criteria of relevance and brevity; while only one of the seven without rule training met both of these criteria.

Discussion

Students given training in the rules of summary writing appear to write better summaries and achieve higher scores on a reading comprehension test than students who are not given such training. This is true at least for the conditions of the tryout and with recognition that evidence was not collected regarding the degree to which pupils have generated the strategy to the point of using it on their own,

applying it to similar learning situations. Asking students to write summaries according to given rules without first training them in the meaning of the rules appears to be detrimental.

Currently, the study is being replicated in eight other schools. Teachers in these schools say that information about pupils' ability to summarize and apply summarization rules is new and important to them. Only one of the six rules - identifying topic sentences - is commonly taught in the present curriculum. These teachers believe that the six skills will contribute to improving comprehension. They indicate that they will continue to teach the rules although two of the teachers said they did not want to generate additional practice exercises, preferring that all exercises be available in textbooks.

Conclusion

Research in reading comprehension is bringing new dimensions to understanding the process by which one comprehends text. This knowledge implies that we should change in the ways we teach and assess reading.

Present achievement tests do not provide the teacher with the information needed for helping pupils improve in their reading comprehension. For example, we do not need so many tests to see if the reader can find the main idea, draw conclusions, and determine cause and effect. Instead we need tests to reveal which of the essential processes or strategies for finding the main idea are known and used by the reader.

Similarly, the materials for teaching reading comprehension should do more than offer practice in finding main ideas, drawing conclusions

and determining cause and effect. They should provide the uninitiated with techniques for achieving these tasks.

References

Brown, A. L., Campione, J. C., and Day, J. D. Learning to learn: On training students to learn from texts. Educational Researcher, Feb. 1981, 14-21.

California State Department of Education. Student Achievement in California Schools. 1979-80 Annual Report, Sacramento, 1980.

Day, J. D. Training summarization skills: A comparison of teaching methods. Unpublished doctoral dissertation. University of Illinois, 1980.

Doctorow, M. J., Wittrock, M. C., Marks, C. B. Generative processes in reading comprehension. Journal of Educational Psychology, 1978, 70, 109-118.

Kintsch, W. and van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85(5), 363-394.

Footnotes

1. This paper was funded by a California Policy Seminar grant. The grant is administered by the Institute of Governmental Services under the direction of Mr. John Cummins. The grant is to Harry Singer (University of California, Riverside), John McNeil and Merl Wittrock (University of California, Los Angeles), and Robert Ruddell (University of California, Berkeley) to study "Achievement in Testing Basic Skills."

School and Classroom Evaluation

M. C. Wittrock & Helen J. Schultze

Purpose of Evaluation

In earlier sections of this report we found that legislators and school leaders require information about student achievement appropriate for the decisions about education that they make for the state and its school districts. Especially useful for these decisions are data about student achievement as it depends upon the type, quality, and length of instruction, the socioeconomic characteristics of the school districts and communities, and most importantly, the intellectual and motivational qualities of the learner. These data include the effects of federally sponsored intervention programs, state aid to local school districts, and the impact on learning of state regulations that affect students and teachers.

At the school and classroom level, we also test students' achievement primarily to evaluate and to improve learning and instruction, and secondarily to evaluate the learners. In each instance we want to know how well, why, and under what conditions students learn or fail to learn.

School and Classroom Uses of Evaluation Data

At the school and classroom levels of evaluation, the measurement of learners, and the measurement of the growth, development, and education of students, which includes more than mastery of basic skills, involves procedures and questions different from those found at the

state and the district levels. In the school and in the classroom, the improvement of instruction involves data about individual learners. Group averages, correlations between abstract factors, and abstract relations between groups of learners and types of teaching should be supplemented with specific information about individual learners. Teachers and administrators need valid information useful for diagnosing learning problems and for prescribing classroom instruction for each learner and for each classroom of learners. These educators need specific information about each student's skill, concepts, knowledge, and comprehension. They need to know as much as possible about the types of instruction or educational programs best for the students. The educators need to know the effects of the teaching procedures, books, and activities used in school. In brief, they need to know far more than the learner's percentile rank on a test that may be only loosely related to the instruction their students received. For these reasons, in the empirical study described in a following section of this paper, we will obtain data on the types of test information elementary school teachers want and need to make teaching decisions. Our goal is to find new and improved ways to provide useful evaluative information to teachers.

One conventional, widespread way to evaluate learning involves purchasing and using commercially published standardized achievement tests that report mainly percentile rankings among students. Even if the skill, information, and comprehension measured by these tests overlap with the school's and the teachers' objectives and curricula, these examinations report only relative standing among students. These tests

do not indicate what or how much any student learned during a year. Nor do these tests help one to discern changes needed in teaching, student motivation and effort, and course content.

Percentile ranks sometimes lead to confusion. An increase in achievement for some students leads to a decrease in their percentile ranks. The reason is that annually normed tests reflect relative achievement among students, not actual achievement.

The Southern California School District Superintendents' Research Council reported that one school district showed an increase from one year to the next in the percentage of correct answers on an achievement test, yet the district's percentile rank dropped. Actually this finding has its positive side, it indicates that other school districts showed even larger gains than the reported school district. Put differently, let us assume that every school district shows some increase in the percentage of answers correct on an achievement test that is renormed every year. With percentile ranks, only those districts that increased more than the average increase for all districts will show an improvement in percentile rank. Those districts which nonetheless showed an increase in student achievement, albeit a less than average increase, will drop in percentile ranking.

It is not only possible but common for students to improve their scores on an achievement test and for them to drop in percentile rankings. These types of data, percentile rankings, are useful for selecting students, for discriminating among them, and for comparing them with one another. However, as they are currently reported, these data leave something to be desired for many of the major decisions teachers and principals make about instruction. In the study described later, we

will examine some of the effects upon teachers' decisions produced by presenting identical reading achievement raw score data in two different formats: norm referenced data and criterion referenced data.

Teachers Uses of Evaluation Data

In sum, we suggest that teachers and educators need test data that meet the following specifications:

1. To evaluate and to improve instruction, teachers need tests and appropriate scoring systems where it is possible for all students to increase their scores, or for all students to decrease their scores, depending upon their actual learning. As their results are now reported in percentile ranks, currently used standardized tests of achievement cannot provide these data.

2. Teachers also need information about individual learners. Teachers and administrators need these data to prescribe or to change instruction. Matrix sampling and related techniques which give a few items to each of a sample of students save student time and save state and district money. These techniques provide useful information for the state legislators and for school district personnel. But these techniques provide little or no information about individual learners that a teacher can use to make decisions about the education best for any one child.

3. Clinical uses, that is, uses of tests for decisions about single individuals, require detailed, valid, reliable, and analytic information about the learners. Reliability that may be high enough for research purposes, e.g., .80, or for working with groups may not be high enough for making recommendations about individuals. The validity of the test items and their overlap or relation to the instruction given in the school weighs heavily in the value of an achievement test to be used for making decisions about individual learners and their future instruction.

4. Test data useful at the school and classroom level should often relate directly to the school's or the teachers' objectives and instruction. We should test learners on what they have studied. Nationally distributed standardized tests have difficulty meeting this criterion in our state and nation, where local control of education is highly valued.

When percentile rankings are highly valued by parents, educational administrators, and legislators, teachers sometimes teach for the standardized test. In this case the test makers, not the educators and the local community, largely determine the curricula and the instruction.

5. At the same time, useful test data sometimes will not be specific to the teacher's or the local school's objectives. As a state, we need to know about the intellectual, moral, and social development of our students. Tests of achievement that consist of locally prescribed content and skills will not tell us about the socially important areas of achievement that have been underemphasized by a school or a teacher. Again, we find reasons to test for multiple types of learning and development that sometimes go beyond a local district's objectives.

6. Test data should provide analytic information about the subskills, the parts, of educational achievement. This characteristic of test data relates to the teacher's job of designing instruction. To know where to focus their time with a learner, a teacher needs reliable test data that indicate the individual learner's strengths and weaknesses, or knowledge and skill in a specific domain. IQ scores and percentile ranks, for example, tell teachers almost nothing about where to focus instruction. If a child cannot read well, is it because of a problem in decoding or comprehension, or perhaps in attention and motivation? Test data cannot

always answer these questions. But analytic data sometimes help answer them better than do the global measures reported as percentile ranks.

7. Test data should be understandable to the people that use them. Most parents, students, and many of the public understand well the meaning of percent of correct answers. They understand less well the meaning of percentile ranks. It is difficult to explain that good teaching or even greatly improved student effort and student learning will not increase percentile rankings at all. It is difficult to explain that any and all funds expended into education, even if used with the greatest efficiency and to the highest effectiveness, will not improve percentile rankings. So long as a state test is renormed each year, at the statewide level, nothing will increase or decrease the percentile rankings of all of the state's schools.

An analogy is a group of people running a footrace around the track. So long as we score them in relation to one another, only one of them can be first, one second, and so on. Good coaching and diligent practice can increase their speed and can cause some of them to pass others of them. But in the end, only one can be first, another must be last. Measuring the number of people who are first and last is a poor way to decide if we have fast runners and good coaches, much less whether we have invested our funds wisely in the coaches, shoes, practice, uniforms, food, and medical care. Rather than fire the coach or question the performances of the runners, we might ask ourselves if we are testing their performance in a way that makes it possible for us to measure learning and achievement. We should measure performance in terms that will allow every learner to improve or to decline regardless of the achievement of the other learners.

8. Test data should help school people to decide among alternative teaching procedures for individual learners. This objective for tests emphasizes the relation between the test data and the characteristics of the learners and the teaching. Simply stated, the data should help you to know with whom and why teaching works well or poorly. It is not enough that tests clearly indicate improvement in student achievement. We need to understand why and now the improvement occurred. Although test data cannot answer these difficult questions, we can design tests and analyze their results in ways that will evidence the relations between the instruction and its effects upon the learners. A closely related problem occurs at the state level when we evaluate intervention programs. Again, the relation between the program, its effects upon students' thoughts, and their effects upon achievement comprise the essence of the evaluation problem. Wittrock (1970) discusses this type of cause-and-effect evaluation, and offers a model for use in educational evaluation.

So far we have discussed teacher needs in the evaluation of learning and the evaluation of instruction. In neither case did we find need for data summarized with percentile ranks, or by comparisons among learners.

9. However, teachers encounter instances when students should be compared or ranked with one another. These instances involve the selection of some students from among many students, such as entrance into special programs, colleges, graduate schools, and professional programs. In these situations, tests designed to discriminate among students according to ability or achievement play important roles in determining admissions. To make these decisions, people do not usually use information about

percentage of items correct, or even information about why learners do well or poorly. Instead, the information they use pertains to the student's performance in relation to that of other students. For these decisions, currently used standardized tests are appropriate. They were designed for these decisions.

Policy Changes

For nearly all of the decisions at the school and classroom level we need a new type of test, in which student learning strategies are identified, and we need to return to using an old type of reporting of test results--one where each student gets a percentage of items correct, one where every student could get the same score as every other student, and one where every student can show an improvement or a decline in learning. We also need analytic tests that are valid, that measure the results of the instruction given to learners, and that reflect the learners' active involvement in the instruction. These types of tests will not be designed to help make decisions about college admissions, or about schools or districts that rank better or worse than other schools or districts. But these tests will be designed to help individual learners to profit from instruction. These tests will be designed to assist parents, teachers, and school administrators to make informed decisions about instruction and its improvement.

An Empirical Study of Test Data and Their Relations To Teaching Decisions

It is not possible in one empirical study to examine all of the issues we have introduced and discussed. Instead, it is more useful and worthwhile to focus the study upon some of the centrally important issues of teaching decisions and test data. One of these issues is the type of test data most useful for the teaching decisions frequently made by classroom teachers. Another issue is the utility of test data about

student learning processes and strategies which lead to, or cause, student achievement. A third issue is the effect upon teachers' decisions about student achievement produced by different formats used to report identical score data, such as criterion referenced tests and norm referenced tests.

To study these three issues we wrote a four part questionnaire. It was revised in a pilot study, and then given to a state-wide sample of elementary school teachers. The questionnaire, the sample, and the procedure are described in the following method section. A copy of the questionnaire appears in Appendix A.

Method

Instrument

A questionnaire (Appendix A) for elementary school teachers was developed to obtain information regarding a) the type of test data most useful for the decisions teachers make regarding their students, b) the utility of test data about student learning strategies, and c) the effects of different reporting formats upon the decisions teachers make about student achievement. The sections of the questionnaire were designed to gather the following information:

1. The first section of the questionnare contained multiple choice questions related to teachers' use of criterion-referenced test (CRT) and norm-referenced test (NRT) information in five major areas: individual student performance, class performance, school performance, differences among learners, and instructional program data.

2. A second section of the questionnaire consisted of multiple choice questions regarding the importance and usefulness to teachers of test information about students' reading comprehension strategies.

3. A third section of the questionnaire was designed to investigate whether teachers made different decisions and evaluations about their students and teaching when presented with CRT information as compared to when presented with NRT information. In this section of the questionnaire teachers compared year-end test reports in reading for two pairs of hypothetical students in the fifth grade: two girls (Mary and Judy) and two boys (Danny and Jeff). The reading test scores for each pair of students represented identical raw scores from subtests of the California Test of Basic Skills (CTBS), Form T, Level 2. These raw scores were converted into percent correct scores (CRT) for one member of each pair (Mary and Danny) and into percentile ranks (NRT) for the other member of each pair (Judy and Jeff). The raw scores for the subtests of Vocabulary, Decoding, Critical Comprehension, and Interpretative Comprehension in the hypothetical reading test reports corresponded, respectively, to raw scores in the CTBS subtests: Reading Vocabulary, Language Mechanics, Reading Comprehension, and Language Expression. The CTBS Examiner's Manual provided the total items in each subtest and the percentile rank for each raw score. The percentile rankings for the NRT format in this study were taken directly from the scales presented in the CTBS. The percent correct score for the CRT format was derived by dividing the raw score by the total items in each subtest. In addition to representing CRT data vs. NRT data, each pair of students also represented an ability grouping: the girls reflected below average ability in their reading scores, whereas the boys represented average ability.

Using the student reading test reports, teachers were asked to compare the two members of each pair in two ways (Mary, CRT, vs. Judy, NRT;

Danny, CRT, vs. Jeff, NRT). Teachers were asked to assign letter grades (A,B,C,D, or F) to each student in Vocabulary and in overall Reading. Values were assigned to each letter grade (A=1, B=2, C=3, D=4, F=5). The mean was calculated and the significance for each group (Mary vs. Judy; Danny vs. Jeff) was determined by a t Test for paired data. A pilot study had revealed no teacher sex bias in student evaluations as a mediating factor in assigning letter grades in vocabulary and in reading.

Teachers were also asked to compare the children of each pair by circling either a YES or a NO response to each of three questions: "Is remedial instruction indicated in Interpretative Comprehension?" "Is this student ready for sixth grade reading?" "Would you promote this student to the next grade?"

4. A fourth section of the questionnaire consisted of questions pertaining to the specific CRT and NRT material used in the teacher's respective school. This section required short answer responses. The data from this section were not analyzed and are not summarized in this report. The variety and idiosyncratic nature of the responses and a failure of many of the participants to answer these questions rendered this section difficult to quantify or to summarize.

The questionnaire was tested in a pilot study with eight teachers in a school in Southern California. The questionnaire consisted of eleven pages: four pages with multiple choice questions pertaining to teacher use of test information and reading strategies; four pages presenting reading test reports for the four hypothetical students; two pages of specific questions pertaining to these student reports; and one page concerning current CRT and NRT use in the teacher's school.

Participants and Procedure

The subjects for this study consisted of a state-wide sample of 43 third and sixth grade teachers who responded to the questionnaire. The study was limited to data collected from third and sixth grade teachers because these teachers have the heaviest testing schedule of all the elementary grades.

Twelve schools from different cities in California were selected for this study. These schools had been used previously in an evaluation study by Drs. Robert Ruddell and Charles Kinzer in the spring of 1981. We thought having the same data base would help us to relate findings across these two studies.

Letters to the principals of these schools explaining the research study and followed by telephone calls resulted in the apparent participation of 65 third and sixth grade teachers. Subsequently, one principal chose not to participate (one questionnaire); two teachers transferred out of their schools and districts; and three completed questionnaires, comprising the entire sample from one school, were lost in return mail. Thus, of the remaining 59 questionnaires, 43 were returned (73% response) representing ten schools.

Results and Discussion

The data from the questionnaire are summarized in the 15 tables presented in the following paragraphs. Most of these tables, all except 14 and 15, report the percentage of the 43 teachers who chose each answer appearing in the questionnaire. Tables 14 and 15 report differences in average letter grades assigned to students as a result of having identical raw scores summarized in two different formats, criterion referenced data and norm referenced data.

Table 1 summarizes the data about the types of test information teachers consider important to them. It was encouraging to find that over 60% of the teachers consider test data important in all of the five categories we examined: Individual student performance, class performance, school performance, differences in achievement among learners, and instructional programs. Between 75% and 80% of the teachers considered test data about class performance and instructional programs to be important to their teaching decisions, and 85% of the teachers reported data about individualized student performance to be important to their work.

Although test data are often criticized for their methodological shortcomings, in this sample teachers indicated strong support for the importance of the five types of test data tested in Table 1, at least for their utility in making teaching decisions.

Next, we examined how these five important types of test data are used. What teaching decisions are based upon them? Table 2 reports the answer to this question. The five types of test data are all used by teachers for planning instruction. On the other hand, only a small percentage of the teachers used any of the five types for assigning grades. A relatively small percentage of the teachers used any of the five types for holding students accountable for learning.

Each of the five types of test information has its own distinctive pattern of utility for teaching decisions. Individual student performance data were reported to be the most generally useful type. It was used by over 50% of the teachers for planning instruction, placing students in special programs, grouping students in class, monitoring students' progress, counseling students, and for making diagnoses. Data about

Table 1

Percentages of Teachers Who Rate Test Information to Be Important To Teaching Decisions

	Types of Test Information				
Importance To Teachers	Individual Student Performance	Class Performance	School Performance	Differences In Achievement Among Learners	Instructional Program
Extremely Important	14.0	4.7	2.3	7.0	9.3
Important	69.8	69.8	60.5	60.5	69.8
Slightly Important	14.0	16.3	30.2	15.3	7.0
Unimportant	0.0	9.3	7.0	7.0	9.3
NO RESPONSE	2.2	0.0	0.0	10.2	4.6

differences in achievement among learners were used by over 50% of the teachers for these same decisions, except for monitoring progress and counseling students. School performance data were used by 50% or more of the teachers only for planning instruction and for comparing schools. Class performance data were used for planning instruction, within class grouping, and diagnosis. Instructional program data were used for planning instruction and within-class grouping. None of the five types of measures was used by 50% or more of the teachers to assign grades or to hold students accountable for learning. Teachers apparently use their own judgments and observations to make these important decisions. Perhaps they do not emphasize student accountability. In any event, test data should be, but are not now, useful to many of these teachers for these two teaching decisions.

Table 2

Percentages of Teachers Who Report Using Test Information
For Making Teaching Decisions

	Types of Test Information				
Teaching Decisions	Individual Student Performance	Class Peformance	School Performance	Differences In Achievement Among Learners	Instructional Program
Planning Instruction	72.1	74.4	55.8	65.1	86.0
Referral/Placement Special Program	86.0	44.2	32.6	55.8	46.5
Within-Class Grouping and Individual Placement	83.7	53.5	25.6	65.1	55.8
Holding Students Accountable for Work, Discipline	27.9	20.9	16.3	30.2	27.9
Assigning Grades	20.9	18.6	9.3	16.3	25.6
Monitoring Students' Progress	62.8	39.5	34.9	39.5	44.2
Counseling and Guiding Students	51.2	30.2	20.9	46.5	18.6
Comparing Groups of Students, Schools, Etc.	34.9	39.5	58.1	25.6	25.6
Diagnostic Procedures	60.5	51.2	30.2	51.2	46.5

Tables 3, 4, and 5 report the percentage of teachers who want test information by item, subtest, and whole test, respectively for the same teaching decisions listed in Table 2. Again, all five types of data are wanted by item, by subtest, and by whole test relatively frequently for planning instruction. Again, individual student performance data are frequently wanted by item and by subtest for most of the teaching decisions listed in the tables. The notable exceptions are holding students accountable, assigning grades, and comparing groups of students and schools.

In fact, none of the five types of test data are wanted frequently, by item, subtest, or whole test, for holding students accountable or for assigning grades. In brief, teachers want item information mainly for planning instruction and diagnoses. They want subtest information mostly for planning instruction and for grouping within class. They want whole test information mainly for planning instruction, placement of students in the special programs, and comparison of groups of students and schools.

Table 3

Percentages of Teachers Who Want Item Test Information For Making Teaching Decisions

	Information By Item				
Teaching Decisions	Individual Student Performance	Class Performance	School Performance	Differences Achievement Among Learners	Instructional Program
Planning Instruction	30.2	37.2	20.9	34.9	37.2
Referra./Placement: Special Program	18.6	9.3	7.0	18.6	4.7
Within-Class Grouping and Individual Placement	20.9	7.0	4.7	18.6	7.0
Holding Students Accountable for Work, Discipline	2.3	2.3	4.7	4.7	4.7
Assigning Grades	7.0	2.3	0.0	2.3	2.3
Monitoring Students' Progress	25.6	18.6	11.6	14.0	14.0
Counseling and Guiding Students	11.6	7.0	2.3	20.9	2.3
Comparing Groups of Students, Schools, Etc.	2.3	4.7	14.0	2.3	4.7
Diagnostic Procedures	34.9	23.3	9.3	23.3	20.9

Table 4

Percentages of Teachers Who Want Subtest Information For Making Teaching Decisions

	Information by Subtest				
Teaching Decisions	Individual Student Performance	Class Performance	School Performance	Differences In Achievement Among Learners	Instructional Program
Planning Instruction	23.3	16.3	16.3	14.0	34.9
Referral/Placement: Special Program	27.9	4.7	4.7	14.0	11.6
Within-Class Grouping for Individual Placement	30.2	20.9	2.3	23.3	27.9
Holding Students Accountable For Work, Discipline	0.0	0.0	2.3	2.3	2.3
Assigning Grades	0.0	0.0	0.0	0.0	4.7
Monitoring Students' Progress	23.3	9.3	0.0	9.3	16.3
Counseling and Guiding Students	11.6	9.3	0.0	16.3	4.7
Comparing Groups of Students, Schools, Etc.	7.0	9.3	16.8	11.6	4.7
Diagnostic Procedures	16.3	11.6	7.0	11.6	11.6

Table 5

Percentages of Teachers Who Want Whole Test Information For Making Teaching Decisions

	Information By Whole Test				
Teaching Decisions	Individual Student Performance	Class Performance	School Performance	Differences In Achievement Among Learners	Instructional Program
Planning Instruction	20.9	18.6	18.6	11.6	32.6
Referral/Placement Special Program	34.9	18.6	9.3	16.3	20.9
Within-Class Grouping And Individual Placement	18.6	9.3	7.0	9.3	16.3
Holding Students Accountable for Work, Discipline	7.0	2.3	7.0	11.6	7.0
Assigning Grades	11.6	9.3	4.7	7.0	7.0
Monitoring Students' Progress	18.6	9.3	11.6	11.6	11.6
Counseling and Guiding Students	7.0	4.7	2.3	14.0	4.7
Comparing Groups of Students, Schools, Etc.	20.9	23.3	27.9	11.6	16.3
Diagnostic Procedures	16.3	14.0	4.7	16.3	14.0

When one juxtaposes the data of Table one with the data of Tables 3, 4 and 5, a curious finding emerges. In Table 1, the teachers indicated the importance of all five types of test data. In Tables 3, 4 and 5, however, they frequently indicated that test data were wanted for decisions involving the planning of instruction, but not for the problems involving grading or holding students accountable. Test data are wanted by the teachers because of their contributions to certain teaching decisions; but, test data are not wanted, or are not seen as useful, for the other teaching decisions listed in these tables.

In Tables 6 and 7 we reported findings about types of test data not often available to elementary school teachers. The data deal with reading comprehension strategies, not with reading achievement,

Table 6

Teachers' Ratings of the Importance of
Test Data About Reading Comprehension Strategies

Teacher Rating	Percentages of Teachers
Extremely Important	39.5
Important	48.8
Slightly Important	2.3
Unimportant	2.3
No Response	7.1

Table 7

Percentages of Teachers Who Want Test Information About Reading Strategies

Reading Strategies	% of Teachers
Obtaining the main idea	90.7
Drawing inferences	86.0
Ability to analyze	81.4
Generalizing from the text	72.1
Ability to synthesize	69.8
Organizing information in the text	69.8
Reading critically	69.8
Constructing summaries	65.1
Generating questions about the text	58.1
Relating text to prior knowledge	53.5
Constructing paragraph headings	51.3
Relating parts of the text to one another	41.9
Constructing meaning and learning from pictures in the text	41.7

Comprehension, vocabulary, or decoding. A large percentage of the teachers, 89%, reported that they rated these data important or extremely important, including 40% of the entire group of teachers who rated data about reading comprehension strategies to be extremely important to their teaching decisions.

Table 7 pursues this finding further. The teachers were asked which reading comprehension strategies they would like to have measured on tests. Nearly all of the strategies, 11 of 13, were selected by 50% or more of the teachers. Several of the strategies, "obtaining the main idea," "drawing inferences," and "ability to analyze" were chosen by over 80% of the teachers. Obviously, test data about comprehension strategies, which mediate achievement, are wanted and valued by the teachers in this sample. However, most modern standardized tests do not provide these data.

It is also interesting to note that the teachers chose as most wanted the comprehension strategies usually stressed in basal reading series--obtaining the main idea and drawing inferences, but did not as frequently choose strategies reading researchers consider critical to comprehension, such as "constructing summaries" and relating the text to prior knowledge."

These differences in preferences among comprehension strategies are a secondary issue. The primary finding is that 88% of the teachers in this study indicated that it was important or extremely important for them to have test data about the reading comprehension strategies of their students. These findings imply that process oriented measures of reading strategies might provide a wanted and useful set of test data for reading teachers.

Table 8

Percentages of Teachers Who Prefer Norm-Referenced Test Data Or Criterion-Referenced Test Data For Teaching Decisions

Teacher Test Information Preference	Types of Test Information				
	Individual Student Performance	Class Performance	School Performance	Differences In Achievement Among Learners	Instructional Program
Norm-Referenced	23.3	25.6	60.5	23.3	34.9
Criterion-Referenced	55.8	51.2	25.6	48.8	44.2
No Response	20.9	23.2	13.9	27.9	20.9

The next and last main part of the questionnaire focused upon the teachers' use of criterion-referenced and norm-referenced test data for making decisions about teaching. Table 8 reports teacher preferences for these two types of data to measure individual, class, and school performance as well as differences in learner achievement and the effectiveness of the instructional programs. For three of these five types of measures, individuals, classes, and differences among students, the teachers chose criterion-referenced data over norm-referenced data by a ratio of two to one. For measuring the effects of the instructional program, the difference in preference was smaller, about 10% in favor of criterion-referenced tests. However, for measuring school performance, the teachers preferred norm-referenced measures over criterion-referenced measures by a ratio of 2 to 1. These data again support the perceived utility of criterion-referenced tests for decisions involving individuals, instructional programs, and, surprisingly, differences in achievement among individual learners. However, for evaluating schools, teachers appropriately preferred norm-referenced data, perhaps because of the problems of constructing and using identical criterion-referenced tests across several schools.

Table 9 shows the percentage of teachers who report using criterion-referenced data or norm-referenced data for teaching decisions regarding below average and average ability students. The patterns of preferences for average and below-average ability students are nearly identical to each other. The only exception is the last item, explaining student performance to parents, where criterion-referenced tests were preferred only for average ability students.

Table 9

Percentages of Teachers Who Report Using Criterion-Referenced Or Norm-Referenced Test Data To Make Teaching Decisions

Teaching Decisions	For Below-Average Student Achievement		For Average Student Achievement	
	CRT	NRT	CRT	NRT
The Reading Abilities/Skills of the Student	46.5	20.9	44.2	14.0
The Reading Level of the Student	18.6	44.2	18.6	32.6
The Kind of Reading Instruction the Student Needs	51.2	18.6	44.2	11.6
The Kind of Materials to Use in Teaching the Student	25.6	18.6	27.9	9.6
Whether the Student is Progressing in Reading	18.6	30.2	16.3	23.3
How to Group the Students for Instruction	18.6	32.6	16.3	23.3
How This Student Compares in Reading to Classmates	11.6	53.5	4.7	48.8
Reading Skills in Which the Student is Experiencing Difficulty	58.1	7.0	48.8	9.3
Whether the Student Should Attempt more Difficult Reading Materials	32.6	18.6	30.2	16.3
How Effective My Reading Program Has Been	18.6	27.9	16.3	25.6
Explaining the Student's Reading Performance to Parents	27.9	30.2	32.6	20.9

For the other items, criterion-referenced tests were used more frequently than norm-referenced tests for decisions about reading skills, reading instruction, reading materials, difficulty of materials, and reading skills. Norm-referenced tests were used more frequently for decisions regarding levels of students, student progress, grouping of students, comparing students to one another, and evaluating program effectiveness.

Tables 10 through 15 report the differential effect upon specific teaching decisions produced by the difference in reporting identical raw scores using criterion-referenced data or norm-referenced data. As we explained earlier in the method section, pages 6 and 7 and 9 and 10 of the questionnaire reported criterion-referenced or norm-referenced data each derived from identical raw scores with fictitious student names added to each set of scores. We asked teachers to make several decisions about the students, including whether: 1) the student needed remedial instruction, 2) the student was ready to read at the next grade level, 3) the student should be promoted, and 4) the letter grade they would assign to the student. By using the same raw scores and by reporting them in two different ways, the differences in decisions made by the teachers should be attributable to the differences in interpretations induced by the reporting systems, not by the performances of the hypothetical statements.

The data for the two students of below-average ability, pages 6 and 7 of the questionnaire, appear in Table 10. Approximately the same number of teachers, 67% and 77%, said that the students needed remedial instruction. However, using criterion-referenced data, 34% of the teachers indicated that the student was ready to read at the next grade

Table 10

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon Teachers' Decisions Regarding Below Average Ability Students Achieving Identical Raw Scores on Standardized Reading Tests

Teaching Decisions	Percentages of Teachers CRT Data		NRT Data	
	Yes	No	Yes	No
Student Needs Remedial Instruction	66.7	33.3	77.1	22.9
Student is Ready for Next Reading Grade Level	34.4	65.6	3.1	96.9
Student Should Be Promoted	79.3	20.7	58.3	41.7

level; while with norm-referenced data, 3% of the teachers thought the student was ready for reading at the next grade level ($p < .004$, binomial test, df=1. For criterion-referenced data and norm-referenced data, respectively, 79.3% and 58.3% of the teachers thought the students should be promoted. A binomial test indicated that this difference approached statistical significance $p < .06$, binomial test, df=1).

For the students of average ability, the differences in teachers' responses to the two different formats, criterion-referenced data and norm-referenced data, are reported in Table 11. These differences are not as large as their counterparts for the students of below-average ability. None of these differences is statistically significant by binomial tests. However, the difference regarding student readiness for sixth grade reading approached statistical significance. ($p < .06$, binomial test, df=1).

However, the letter grades teachers give to the same raw scores, reported differently, did differ statistically from each other by t-test for matched groups, for the below-average students ($p < .001$) and for the average students ($p < .002$). The letter grades teachers assigned for vocabulary skills and overall reading scores are presented in Table 12 and 13 for the below-average students. The criterion-referenced data resulted in a higher letter grade for vocabulary ($p < .001$) and for overall reading ($p < .009$) than did the norm-referenced data, again using t-tests for matched groups.

Tables 14 and 15 present the comparable data for students of average ability. Again the criterion-referenced data produced a statistically significant higher mean letter grade than did the norm-referenced data for vocabulary ($p < .002$) and for overall reading score ($p < .008$),

Table 11

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon Teachers' Decisions Regarding Average Ability Students Who Achieved Identical Raw Scores on Standardized Reading Tests

Teaching Decisions	Percentages of Teachers			
	CRT Data		NRT Data	
	Yes	No	Yes	No
Student Needs Remedial Instruction	37.5	62.5	29.0	71.0
Student is Ready for Next Reading Grade Level	94.1	5.9	80.0	20.0
Student Should Be Promoted	97.1	2.9	100.0	0.0

Table 12

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon The Letter Grades Teachers Gave Below-Average Ability Students Who Achieved Identical Raw Scores on Standardized Tests of Vocabulary and Reading

Percentages of Teachers Assigning Each Letter Grade				
	Vocabulary Skills		Reading Grade Overall	
Letter Grade	CRT	NRT	CRT	NRT
A	0.0	0.0	0.0	0.0
B	8.8	3.1	0.0	0.0
C	38.2	18.8	48.4	21.3
D	52.9	56.3	42.5	63.6
F	0.0	21.9	9.1	15.2

Table 13

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon the Grades Teachers Assigned Below-Averaged Ability Students Who Achieved Identical Raw Scores on a Standardized Reading Test

Reading Skills		CRT	NRT	p*
Vocabulary Skills	Mean	3.37	3.96	0.001
	s.d.	0.69	0.76	
Reading Grade Overall	Mean	3.56	3.96	
	s.d.	0.70	0.65	0.009

*p value for differences between means determined by Student's t-test; df = 26.

Numerical equivalents of letter grades:

A = 1, B = 2, C = 3, D = 4, & F = 5.

Table 14

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon The Letter Grades Teachers Gave Average Ability Students Who Achieved Identical Raw Scores on Standardized Tests of Vocabulary and Reading

Percentages of Teachers Assigning Each Letter Grade				
	Vocabulary Skills		Reading Grade Overall	
Letter Grade	CRT	NRT	CRT	NRT
A	0.0	0.0	0.0	0.0
B	34.4	16.7	29.0	13.3
C	62.5	56.7	61.3	60.0
D	3.1	16.7	9.7	16.7
F	0.0	10.0	0.0	10.0

Table 15

Effects of Criterion-Referenced Data vs. Norm-Referenced Data Upon the Grades Teachers Assigned Average Ability Students Who Achieved Identical Raw Scores on a Standardized Readig Test

Reading Skills		CRT	NRT	p*
Vocabulary Skills	Mean	2.67	3.19	0.002
	s.d.	0.56	0.88	
Reading Grade Overall	Mean	2.78	3.22	
	s.d.	0.58	0.85	0.009

*p value for differences between means determined by Student's t-test; df = 26.

Numerical equivalents of letter grades:

A = 1, B = 2, C = 3, D = 4, & F = 5.

using t-tests for matched groups. In these two instances, the difference in letter grades due to the difference in the reporting systems was ½ of a letter grade for vocabulary and again for overall reading.

In both of the examples, the methods of reporting the data to teachers produced sizable differences in the grades teachers assigned the same raw scores, for both students of below-average ability and for students of average ability. With other teacher decisions, such as remediation, reading readiness, and student promotion in grade, the reporting systems had less difference in effect. With criterion-referenced data teachers felt the below-average ability student was ready to read at the next grade level more frequently than they felt the same student was ready when the scores were presented in norm-referenced form. With criterion-referenced data teachers showed a trend ($p < .06$) to consider below average students ready for 6th grade reading and also ready for promotion more frequently than they did when the identical raw scores were presented in norm-referenced data. In sum, the form of the data reporting system does influence the teacher with some decisions, including the assignment of letter grades to students.

In the introduction to this study, we hypothesized that teachers want and need new and improved types of data, and old formats for reporting them, to facilitate some of the decisions they make about student achievement. Our questionnaire data support these beliefs, at least with some of the teachers' most important instructional decisions.

Individual student data, not available from some state-wide administered achievement tests, were reported by the teachers in our sample to be the most widely useful source of information for classroom decisions. These data were used by over 50% of the teachers for planning instruction,

placing students in special programs, grouping students within classes, monitoring students' progress, counseling students, and making diagnoses about student learning. On the other hand, none of the five main categories of data was used by 50% or more of the teachers for assigning grades or for holding students accountable for learning.

In the second part of the questionnaire the teachers indicated that data about reading processes, rather than reading achievement, were important to 88% of them and extremely important to 40% of them. These high percentages indicate a possible need for a new type of diagnostically useful process data not currently available in widely used standardized achievement tests. Test information about the strategies children use to read with comprehension would provide teachers with useful data about some of the causes of achievement, and about some of the cognitive areas in need of remedial instruction. Process oriented measures provide a new and potentially important source of information for teachers that leads to hypotheses about the causes of reading comprehension and to the design of instruction to teach strategies of reading comprehension. These process oriented measures might also be valuable in other areas of basic skills.

In the third part of the questionnaire we examined the effects of criterion-referenced data and norm-referenced data upon several teaching decisions. With identical raw scores, we found that teachers usually gave higher letter grades to students when the raw scores were reported in criterion-referenced fashion. This finding occurred with average and below average ability students, the only two groups of hypothetical students that we presented to the teachers. With only one of three other teacher decisions, readiness to read at the next grade, did we

find a statistically signficiant difference favoring criterion-referenced over norm-referenced data for students of below-average ability ($p < .004$), and approaching significance ($p < .06$), for students of average ability. Criterion- referenced data resulted in a larger percentage of teachers (34.4 vs. 3.1) reporting that the below-average student was ready for reading at the next grade level. However, the pattern of differences across all comparisons of criterion-referenced data vs. norm-referenced data indicated, with one exception out of eight, that criterion-referenced data produced more positive evaluations of student performance than did norm-referenced data.

In sum, there is support for the hypotheses we investigated in this study. New types of data about individual students and their learning strategies are wanted and needed by classroom teachers for instructional decision making. The format used to report the data does influence some major teaching decisions, including letter grading, judgments about student readiness to read at the next grade level, and perhaps also promotion to the next grade.

References

Wittrock, M.C. The evaluation of instruction: Cause-and-effect relations in naturalistic data. Ch. in M.C. Wittrock and D.E. Wiley (Eds). The evaluation of instruction: Issues and problems. New York: Holt, 1970, pp. 3-21.

Appendix 1

Questionnaire

This questionnaire is part of a research study conducted by faculty members at UCLA. The purpose of the study is to find ways to improve the quality and utility of test information about reading comprehension available to reading teachers.

We would like to know the types of test information that are most useful to you as a reading teacher. The attached questionnaire is designed to tell us your needs and wishes about information from tests of reading comprehension.

Please complete the questionnaire and return it to the secretary in the main office in your school or in the principal's office. Your responses will be collated with the responses of other reading teachers to give us an understanding of the needs of a group of reading teachers. Your individual responses will be kept confidential. We are interested only in the total group's responses.

You may omit your name from this questionnaire if you wish, but please tell us the grade you teach and the approximate number of students and classes you teach.

Thank you for your cooperation. We hope that this study will help to produce new and better types of test information for reading teachers to use in their daily instructional decisions.

NAME (OPTIONAL) _____

GRADE LEVEL TAUGHT_____

NUMBER OF CLASSES TAUGHT_____

TEACHING EXPERIENCE, YEARS_____

SCHOOL_____

CITY_____

In the following pages the terms norm-referenced and criterion-referenced tests and testing occur frequently. Norm-referenced instruments are used to determine an individual's relative standing in comparison with a norm group (such as in achievement and aptitude tests). Criterion-referenced instruments compare a student to a level of performance he is expected to achieve in a carefully defined domain of behaviors (such as in appropriate citizenship skills, math skills).

KINDS OF INFORMATION AND THEIR USES

1. INDIVIDUAL STUDENT PERFORMANCE

(a)How important are test data about individual student achievement to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b)For what teaching decisions do you use this information?

( )1.Planning Instruction
( )2.Referral/Placement: Selection, Special Program
( )3. Within-Class Grouping and Individual Placement
( )4.Holding Students Accountable for Work, Discipline
( )5.Assigning Grades
( )6.Monitoring Student Progress
( )7.Counseling and Guiding Students
( )8.Comparing Groups of Students, Schools, Etc.
( )9.Diagnostic Procedures
( )10.Other (Write in):_____

(c) For each of the teaching decisions selected above, would you like to have test information by: (Please write in the numbers.)

1.ITEM (_,_,_,_) 2.SUBTEST (_,_,_,_) 3.WHOLE TEST (_,_,_,_)

(d) What kind of test information about students is more important to you?

( )1.NORM-REFERENCED ( )2. CRITERION-REFERENCED

2. CLASS PERFORMANCE

(a) How important are test data about class performance to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b)For what teaching decisions do you use this information?

( )1.Planning Instruction
( )2.Referral/Placement: Selection, Special Program
( )3.Within-Class Grouping and Individual Placement
( )4.Holding Students Accountable for work, Discipline
( )5.Assigning Grades
( )6.Monitoring Student Progress
( )7.Counseling and Guiding Students
( )8.Comparing Groups of Students, Schools, Etc.
( )9.Diagnostic Procedures
( )10.Other (Write in):_____

(c) For each of the teaching decisions selected above, would you like to have test information by: (Please write in the numbers.)

1.ITEM (_,_,_,_) 2.SUBTEST (_,_,_,_) 3.WHOLE TEST (_,_,_,_)

(d) What kind of test information about class performance is more important to you?

( )1.NORM-REFERENCED ( )2.CRITERION-REFERENCED

3. SCHOOL PERFORMANCE

(a) How important are test data about school achievement to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b) For what teaching decisions do you use this information?

( )1.Planning Instruction
( )2.Referral/Placement: Selection, Special Program
( )3.Within-Class Grouping and Individual Placement
( )4.Holding Students Accountable for Work, Discipline
( )5.Assigning Grades
( )6.Monitoring Student Progress
( )7.Counseling and Guiding Students
( )8.Comparing Groups of Students, Schools, etc.
( )9. Diagnostic Procedures
( )10.Other (Write in)_____

(c) For each of the teaching decisions selected above, would you like to have test information by: (Please write in the numbers.)

1.ITEM (_,_,_,_) 2.SUBTEST (_,_,_,_) 3.WHOLE TEST (_,_,_,_)

(d) What kind of information about school achievement is more important to you

( )1.NORM-REFERENCED ( )2.CRITERION-REFERENCED

4. DIFFERENCES IN ACHIEVEMENT AMONG LEARNERS

(a) How important are test data about differences in achievement among learners to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b) For what teaching decisions do you use this information?

( )1.Planning Instruction
( )2.Referral/Placement: Selection, Special Program
( )3.Within-Class Grouping and Individual Placement
( )4.Holding Students Accountable for Work, Discipline
( )5.Assigning Grades
( )6.Monitoring Student Progress
( )7.Counseling and Guiding Students
( )8.Comparing Groups of Students, Schools, etc.
( )9.Diagnostic Procedures
( )10.Other (Write In):_____

(c) For each of the teaching decisions selected above, would you like to have test information by: (Please write in the numbers.)

1.ITEM (_,_,_,_) 2.SUBTEST (_,_,_,_) 3.WHOLE TEST (_,_,_,_)

(d) What kind of test information abouts students' differences in achievement is more important to you?

( )1.NORM-REFERENCED ( )2.CRITERION-REFERENCED

5. INSTRUCTIONAL PROGRAM DATA

(a) How important are test data about the instructional program to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b) For what teaching decisions do you use this information?

( )1.Planning Instruction
( )2.Referral/Placement: Selection, Special Program
( )3.Within-Class Grouping and Individual Placement
( )4.Holding Students Accountable for Work, Discipline
( )5.Assigning Grades
( )6.Monitoring Student Progress
( )7.Counseling and Guiding Students
( )8.Comparing Groups of Students, Schools, etc.
( )9.Diagnostic Procedures
( )10.Other (Write In):_____

(c) For each of the teaching decisions selected above, would you like to have test information by: (Please write in the numbers.)

1.ITEM (_,_,_,_) 2.SUBTEST (_,_,_,_) 3.WHOLE TEST (_,_,_,_)

(d)What kind of information about the instructional program is more important to you?

( )1.NORM-REFERENCED ( )2.CRITERION-REFERENCED

6.STUDENT'S READING COMPREHENSION STRATEGIES

(a) How important are test data about individual student reading comprehension strategies to you?

( )1.Extremely Important ( )2.Important ( )3.Slightly Important ( )4.Unimportant

(b)For which of the following strategies would you like to have test information?

( )1.Ability to draw inferences
( )2.Ability to construct summaries
( )3.Ability to construct paragraph headings
( )4.Ability to get the main idea
( )5.Ability to construct meaning and learn from pictures in the text
( )6.Ability to relate text to their prior knowledge
( )7.Ability to read critically
( )8.Ability to synthesize
( )9.Ability to analyze
( )10.Ability to relate parts of the text to one another
( )11.Ability to generate questions about the text
( )12.Ability to organize information from the text
( )13.Ability to generalize from the text

(c)For which reading comprehension strategies not mentioned above would you like to have test information?

_____

_____

The reading test results for 2 fifth grade students, Mary Smith and Judy Jones, are presented on the next two pages.

Based on the information presented in each of the 2 reports, which report is more useful to you in making the following decisions? Please indicate your response by writing Mary or Judy in front of each number.

A. THIS TEST REPORT HELPS ME DETERMINE:

_____1.The reading abilities/skills of the student.

_____2.The reading level of the student.

_____3.The kind of reading instruction the student needs.

_____4.The kinds of materials to use in teaching the student.

_____5.Whether the student is progressing in reading.

_____6.How to group the student for instruction.

_____7.How this student compares in reading to her classmates.

_____8.The reading skills in which the student is experiencing difficulty.

_____9.Whether the student should attempt more difficult reading materials

_____10.How effective my reading program has been.

_____11.How to explain the student's reading performance to her parents.

B.PLEASE ANSWER THESE SPECIFIC QUESTIONS ABOUT MARY BASING YOUR RESPONSES SOLELY ON HER READING TEST RESULTS. (Circle one.)

1.What letter grade would you assign Mary in Vocabulary Skills? (A,B,C,D,F)
2.Is remedial instruction indicated in Interpretative Comprehension?(Yes or No)
3.What letter grade would you assign Mary in Reading overall? (A,B,C,D,F)
4.Is this student ready for 6th grade reading? (Yes or No)
5.Would you promote Mary to the next grade? (Yes or No)

C.PLEASE ANSWER THESE SPECIFIC QUESTIONS ABOUT JUDY BASING YOUR RESPONSES SOLELY ON HER READING TEST RESULTS. (Circle one.)

1.What letter grade would you assign Judy in Vocabulary Skills? (A,B,C,D,F)
2.Is remedial instruction indicated in Interpretative Comprehension?(Yes or No)
3.What letter grade would you assign Judy in Reading overall? (A,B,C,D,F)
4.Is this student ready for 6th grade reading? (Yes or No)
5.Would you promote Judy to the next grade? (Yes or No)

SCHOOL NO. 4321

STUDENT NAME: Smith, Mary BIRTHDATE: 02/12/70

SPRING 1981 - GRADE 5

COMPREHENSIVE TESTS OF BASIC SKILLS

	TOTAL ITEMS	PERCENT CORRECT
Vocabulary	40	63
Decoding	20	55
Critical Comprehension	45	42
Interpretative Comprehension	35	60

SCHOOL NO. 4321

STUDENT NAME: Jones, Judy BIRTHDATE: 03/11/70

SPRING 1981 - GRADE 5

COMPREHENSIVE TESTS OF BASIC SKILLS

	RAW SCORE	PERCENTILE RANK
Vocabulary	25	34
Decoding	11	36
Critical Comprehension	19	31
Interpretative Comprehension	21	42

The reading test results for 2 fifth grade students, Danny Wilson and Jeff rown, are presented on the next two pages.

Based on the information presented in each of the 2 reports, which report is re useful to you in making the following decisions? Please indicate your response by writing Danny or Jeff in front of each number.

A. THIS TEST REPORT HELPS ME DETERMINE:

_____1.The reading abilities/skills of the students.

_____2.The reading level of the students.

_____3.The kind of reading instruction the student needs.

_____4.The kinds of materials to use in teaching the student.

_____5.Whether the student is progressing in reading.

_____6.How to group the student for instruction.

_____7.How this student compares in reading to her classmates.

_____8.The reading skills in which the student is experiencing difficul

_____9.Whether the student should attempt more difficult reading materials.

_____10.How effective my reading program has been.

_____11.How to explain the student's reading performance to her parents.

B.PLEASE ANSWER THESE SPECIFIC QUESTIONS ABOUT DANNY BASING YOUR RESPONSES SOLELY ON HIS READING TEST RESULTS. (Circle one.)

1.What letter grade would you assign Danny in Vocabulary Skills? (A,B,C,D,F)
2.Is remedial instruction indicated in Interpretative Comprehension?(Yes or N
3.What letter grade would you assign Danny in Reading overall? (A,B,C,D,F)
4.Is this student ready for 6th grade reading? (Yes or No)
5.Would you promote Danny to the next grade? (Yes or No)

C.PLEASE ANSWER THESE SPECIFIC QUESTIONS ABOUT JEFF BASING YOUR RESPONSES SOLELY ON HIS READING TEST RESULTS. (Circle one.)

1.What letter grade would you assign Jeff in Vocabulary Skills? (A,B,C,D,F)
2.Is remedial instruction indicated in Interpretative Comprehension?(Yes or
3.What letter grade would you assign Jeff in Reading overall? (A,B,C,D,F)
4.Is this student ready for 6th grade reading? (Yes or No)
5.Would you promote Jeff to the next grade? (Yes or No).

SCHOOL NO. 4321

STUDENT NAME: Wilson, Danny BIRTHDATE:02/6/70

SPRING 1981 - GRADE 5

COMPREHENSIVE TESTS OF BASIC SKILLS

	TOTAL ITEMS	PERCENT CORRECT
Vocabulary	40	73
Decoding	20	70
Critical Comprehension	45	60
Interpretative Comprehension	35	69

SCHOOL NO. 4321

STUDENT NAME: Brown, Jeff BIRTHDATE: 03/7/70

SPRING 1981 - GRADE 5

COMPREHENSIVE TESTS OF BASIC SKILLS

	RAW SCORE	PERCENTILE RANK
Vocabulary	29	52
Decoding	14	55
Critical Comprehension	27	57
Interpretative Comprehension	24	58

Please answer the following questions concerning the current testing program you use.

A. What published norm-referenced tests are given to your students?

B. And how do you use these data?

C. What criterion-referenced tests are given to your students?

D. And how do you use these data?

E. What tests do you construct and give to your students?

Please indicate whether they are norm- or criterion-referenced tests.

F. What improvements would you most like to see in your current testing program?